



Quantification of Structure from Medical Images

Qazi, Arish Asif

Publication date:
2008

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Qazi, A. A. (2008). *Quantification of Structure from Medical Images*. Department of Computer Science, University of Copenhagen.



Quantification of Structure from Medical Images

By

ARISH ASIF QAZI

Quantification of Structure from Medical Images

Arish Asif Qazi

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
of the
University of Copenhagen



October 2008

Department of Computer Science

Front cover: Reconstruction of the Corticospinal tract (yellow) and the Superior-Longitudinal Fasciculus (green) by streamline tractography. Colored blobs represent regions of fMRI activations.

Quantification of structure from medical images
© Arish Asif Qazi

Department of Computer Science
Image Group
University of Copenhagen
Telephone: + 45 (0)3532 1400

To my wife and my parents

Abstract

In this thesis, we present automated methods that quantify information from medical images; information that is intended to assist and enable clinicians gain a better understanding of the underlying pathology. The first part of the thesis presents methods that analyse the articular cartilage, segmented from MR images of the knee. The cartilage tissue is considered to be a key determinant in the onset of Osteoarthritis (OA), a degenerative joint disease, with no known cure. The primary obstacle has been the dependence on radiography as the ‘gold standard’ for detecting the manifestation of cartilage changes. This is an indirect assessment, since the cartilage is not visible on x-rays. We propose *Cartilage Homogeneity*, quantified from MR images, as a marker for detection of the early biochemical alterations in the articular cartilage. We show that homogeneity provides accuracy, sensitivity, and information beyond that of traditional morphometric measures.

The thesis also proposes a fully automatic and generic statistical framework for identifying biologically interpretable regions of difference (ROD) between two groups of biological objects, attributed by anatomical differences or changes relating to pathology, without *a priori* knowledge about the location, extent, or topology of the ROD. Based on quantifications from both morphometric and textural based imaging markers, our method has identified the most pathological regions in the articular cartilage.

The remaining part of the thesis presents methods based on diffusion tensor imaging, a technique widely used for analysis of the white matter of the central nervous system in the living human brain. An inherent drawback of the traditional diffusion tensor model is its limited ability to provide detailed information about multi-directional fiber architecture within a voxel. This leads to erroneous fiber tractography results in locations where fiber bundles cross each other. We present a novel tractography technique, which successfully traces through regions of crossing fibers. Detection of crossing white matter pathways can improve neurosurgical visualization of functionally relevant white matter areas. We also present preliminary results of analysing the meshwork of the collagen fibers in the articular cartilage by high-resolution diffusion tensor imaging.

Keywords—Osteoarthritis, Imaging Biomarkers, Knee Cartilage, Diffusion Tensor Imaging, Crossing Fibers, Brain.

Acknowledgements

I would first like to thank and express my deep and sincere gratitude to my supervisors, Mads Nielsen and Erik Dam. Mads is an extremely learned, honest, and sincere person; I have learned a lot from him. Thank you so much for the career advice. Erik, well it's difficult to express my gratitude in words but I will give it a try. He has been a constant source of knowledge, inspiration, and support throughout my PhD. But what I have really learnt from him and what I cherish the most is, patience. I remember, sometimes showering him with the stupidest questions and he would never hesitate to answer them. Today, thinking of them makes me laugh. I have never seen him getting annoyed!. Thank you, Erik.

I would like to thank Ole Fogh Olsen for supervising the first six months of my PhD. I am extremely thankful to Carl-Fredrik Westin for supervising me and helping me out while I was at Harvard Medical School. Indeed, it has been a huge learning experience for me. Tons of thanks to Gordon Kindlmann and Lauren O'Donnell. The two-tensor work wouldn't have been possible without them. Thanks to Alexandra Golby for the discussions on neurosurgery and brain anatomy. Thanks to all my friends, and colleagues, mostly at LMI, the image group at DIKU, and Nordic Bioscience. Thanks to Imran Siddiqi for proof-reading my thesis, I am lucky to have a friend like him. I am grateful to CCBR - Synarc and Nordic Bioscience for funding my PhD.

Above all, I would like to thank my family, my parents and my wife. Indeed, the hard work of my parents has made the impossible possible for me. I have not come across a more sincere, honest, and hardworking person than my father. Thank you, Baba and Ammi.

Words are simply not enough to express my feelings and gratitude for Nazish, my wife. She has been the constant source of support throughout my PhD. She has always been there for me; she has taught me the nature of true love. Nazish is a source of happiness and love in my life and without her, this thesis, simply, would not have happened.

Contents

ABSTRACT	I
ACKNOWLEDGEMENTS	II
CONTENTS	III
PREFACE	VII
LIST OF PUBLICATIONS	VIII
1 INTRODUCTION	1
1.1 Purpose	2
1.2 Contributions and Outline	2
I FROM MEDICAL IMAGES TO LOCALIZATION OF ANATOMICAL CHANGES	4
2 OSTEOARTHRITIS: IMAGING AND BIOMARKERS	5
2.1 Osteoarthritis (OA)	5
2.2 Knee Osteoarthritis	8
2.2.1 Imaging	8
2.2.2 Biomarkers	12
3 AUTOMATIC QUANTIFICATION OF CARTILAGE HOMOGENEITY	14
3.1 Introduction	15
3.1.1 Related Work	16
3.2 Methods	16
3.2.1 Imaging	16
3.2.2 Cartilage Homogeneity	17
3.3 Results	18
3.3.1 Manual Segmentations	18
3.3.2 Automatic Segmentations	19
3.3.3 Reproducibility	20
3.4 Conclusion	21

4 SEPARATION OF HEALTHY AND EARLY OSTEOARTHRITIS BY AUTOMATIC QUANTIFICATION OF CARTILAGE HOMOGENEITY	22
4.1 Introduction	24
4.2 Materials and Methods	25
4.2.1 Population	25
4.2.2 Imaging Protocol	25
4.2.3 Automatic Cartilage Segmentation	26
4.2.4 Quantification of Homogeneity	27
4.2.5 Statistical Analysis	29
4.3 Results	29
4.3.1 Cartilage Homogeneity Measure	29
4.3.2 JSW, Cartilage Volume, Intensity, and Entropy Quantifications	29
4.3.3 Quantification Precision	31
4.4 Discussion	31
5 SEPARATION OF HEALTHY AND EARLY OA BY QUANTIFICATION OF CARTILAGE HOMOGENEITY: A LONGITUDINAL STUDY	35
5.1 Introduction	35
5.2 Objective	35
5.3 Methods	35
5.4 Results	36
6 OSTEOARTHRITIC CARTILAGE IS MORE HOMOGENEOUS THAN HEALTHY CARTILAGE - IDENTIFICATION OF A SUPERIOR ROI CO- LOCALISED WITH A MAJOR RISK FACTOR FOR OSTEOARTHRITIS	37
6.1 Introduction	39
6.2 Materials and Methods	40
6.2.1 Population	40
6.2.2 Imaging Protocol	42
6.2.3 Automatic Cartilage Segmentation	42
6.2.4 Quantification of Homogeneity	42
6.2.5 Statistical Analysis	43
6.2.6 Cartilage Region Partitioning	44
6.3 Results	47
6.3.1 Cartilage Volume and Homogeneity	47
6.3.2 Cartilage Partitioning	48
6.4 Discussion	49
7 A VARIATIONAL METHOD FOR AUTOMATIC LOCALIZATION OF THE MOST PATHOLOGICAL ROI IN THE KNEE CARTILAGE	53

7.1	Introduction	54
7.2	Materials and Methods	55
7.2.1	Imaging and Cartilage Segmentation	55
7.2.2	Localizing the ROI	55
7.3	Results	60
7.3.1	Generalization and Anatomical Plausibility	60
7.3.2	Regularization of the ROI	61
7.4	Conclusions	61
8	IDENTIFYING REGIONS OF DIFFERENCE FOR SAMPLE SIZE MINIMIZATION	64
8.1	Introduction	65
8.2	Preliminaries	67
8.3	Methods	68
8.3.1	Regions of difference Identification by Sample Size Optimization (RISSO)	68
8.3.2	Optimization of the Functional	72
8.3.3	Finding the Optimal Regularization Parameters	72
8.3.4	LDA	73
8.4	Results on Synthetic Data	73
8.4.1	Construction	73
8.4.2	Evaluation	74
8.5	Results on Clinical Data	75
8.5.1	Data Characteristics	76
8.5.2	Cartilage Segmentation and Alignment	76
8.5.3	Feature Measurements	76
8.5.4	Intensity Non-uniformity Correction	77
8.5.5	Evaluation	79
8.5.6	Comparison with LDA	81
8.6	Discussion and Conclusions	82
II	ANALYSIS OF COMPLEX ANATOMICAL STRUCTURE BY DIFFUSION MRI	85
9	DIFFUSION TENSOR MAGNETIC RESONANCE IMAGING	86
9.1	White Matter Anatomy	86
9.2	Diffusion MRI	88
9.2.1	Anisotropic Diffusion and the Tensor	90
9.2.2	Tensor Visualization and Tractography	92

10	RESOLVING CROSSINGS IN THE CORTICOSPINAL TRACT BY TWO-TENSOR STREAMLINE TRACTOGRAPHY: METHOD AND CLINICAL ASSESSMENT USING FMRI	96
10.1	Introduction	98
10.2	Materials and Methods	102
10.2.1	Background Theory: Single Tensor Modeling and Tractography	102
10.2.2	Two-Tensor Model	102
10.2.3	eXtended Streamline Tractography (XST)	104
10.2.4	Probabilistic Tractography	105
10.2.5	Synthetic Data Generation	105
10.2.6	Human Brain Data	105
10.2.7	Functional MRI (fMRI) Data	106
10.2.8	Experiments	106
10.3	Results	106
10.3.1	Simulated Data	106
10.3.2	In Vivo Fiber Tractography	107
10.4	Discussion	112
10.5	Conclusion	113
11	TWO-TENSOR DIFFUSION MRI ANALYSIS ENABLES VISUALIZATION OF ARTICULAR CARTILAGE FIBER CROSSINGS	115
11.1	Structure of the Articular Cartilage	115
11.2	Purpose	116
11.3	Methods	117
11.4	Results	119
11.5	Conclusion	120
12	CONCLUSIONS AND PERSPECTIVES	121
	LIST OF FIGURES	124
	LIST OF TABLES	128
	BIBLIOGRAPHY	129

Preface

The work described in this thesis was carried out between November 2005 and October 2008 at the Department of Computer Science, University of Copenhagen, in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Computer Science. The project was funded by CCBR - Syncarc, and Nordic Bioscience Imaging.

The project was supervised by Prof. Dr. Mads Nielsen (University of Copenhagen) and Dr. Erik Dam (Nordic Bioscience Imaging). The first six months of the Ph.D. were supervised by Dr. Ole Fogh Olsen (formerly Associate Professor at the IT University of Copenhagen and currently at Oticon). A significant part of the Ph.D. research was carried out at the Laboratory of Mathematics in Imaging (LMI), Harvard Medical School, USA. The work at LMI was supervised by Prof. Dr. Carl-Fredrik Westin.

The focus of the thesis has been research and development of automated methods for analysis of medical images. In particular, the efforts have been directed to development of: imaging markers from magnetic resonance images that are able to reliably quantify progression of the disease, Osteoarthritis, and methods that aim to localize regions in the articular cartilage, which are most affected by the disease. The dissertation also presents techniques based on diffusion tensor imaging to analyse the complex structure of the brain white matter, and the collagen matrix of the articular cartilage. The presented thesis is an amalgam of selected research papers that have been either accepted or are in review in major medical imaging journals and conferences.

Copenhagen, October 2008

Arish A. Qazi

List of Publications

Papers in International Journals

1. **Arish A. Qazi**, Erik B. Dam, Marco Loog, Mads Nielsen, "*Identifying Regions of Difference for Sample Size Minimization*". [in review]
2. **Arish A. Qazi**, Alireza Radmanesh, Lauren O'Donnell, Gordon Kindlmann, Sharon Peled, Stephan Whalen, Carl-Fredrik Westin, Alexandra J. Golby, "*Resolving Crossings in the Corticospinal Tract by Two-tensor Streamline Tractography: Method and Clinical Assessment using fMRI*", *NeuroImage*, available online July 2008.
3. **Arish A. Qazi**, Jenny Folkesson, Paola Pettersen, Morten A. Karsdal, Claus Christiansen, Erik B. Dam, "*Separation of Healthy and Early Osteoarthritis by Automatic Quantification of Cartilage Homogeneity*", *Osteoarthritis and Cartilage*, Volume 15, issue 10, pages 1199-1206, October 2007.
4. **Arish A. Qazi**, Erik B. Dam, Mads Nielsen, Morten A. Karsdal, Paola Pettersen, Claus Christiansen, "*Osteoarthritic Cartilage is more Homogeneous than Healthy Cartilage – Identification of a Superior ROI Co-localised with a Major Risk Factor for Osteoarthritis*", *Academic Radiology*, Volume 14, issue 10, pages 1209-1220, October 2007.

Papers in Conference Proceedings

5. **Arish A. Qazi**, Gordon Kindlmann, Lauren O'Donnell, Sharon Peled, Alireza Radmanesh, Stephan Whalen, Alexandra J. Golby, Carl-Fredrik Westin, "*Two-Tensor Streamline Tractography through White Matter Intra-voxel Fiber Crossings: Assessed by fMRI*", In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR): MMBIA*, 2008.
6. **Arish A. Qazi**, Erik B. Dam, Marco Loog, Mads Nielsen, Francois Lauze, Claus Christiansen, "*A Variational Method for Automatic Localization of the most Pathological ROI in the Knee Cartilage*", In *Proceedings of SPIE Medical Imaging: Image Processing*, 6914, 2008.

7. **Arish A. Qazi**, Erik B. Dam, Ole F. Olsen, Mads Nielsen, Claus Christiansen, "*Automatic Detection of Diseased Regions in Knee Cartilage*", In Proceedings of SPIE Medical Imaging: Image Processing, 6512, 2007.
8. **Arish A. Qazi**, Ole F. Olsen, Erik B. Dam, Jenny Folkesson, Paola Pettersen, Claus Christiansen, "*Automatic Quantification of Cartilage Homogeneity*", In Proceedings of MICCAI: Quantitative Automated Musculoskeletal Analysis, 2006.

Published Abstracts

9. **Arish A. Qazi**, Erik B. Dam, Sharon Peled, Carl-Fredrik Westin, "*Two-Tensor Diffusion MRI Analysis Enables Visualization of Articular Cartilage Fiber Crossings*", In Proceedings of OARSI, 2008.
10. **Arish A. Qazi**, Gordon Kindlmann, Carl-Fredrik Westin, "*Simulated Diffusion Dataset for Multi-tensor Fiber Tractography*", In Proceedings of Human Brain Mapping, 2008.
11. **Arish A. Qazi**, Morten A. Karsdal, Claus Christiansen, Erik B. Dam, "*Separation of Healthy and Early OA by Quantification of Cartilage Homogeneity: A Longitudinal Study*", In Proceedings of Workshop on Imaging Based Measures of OA, 2007.
12. **Arish A. Qazi**, Erik B. Dam, Morten A. Karsdal, Mads Nielsen, Claus Christiansen, "*Homogeneity Discrimination leads to Identification of a Superior ROI in Knee Cartilage*", In Proceedings of The Annual European Congress of Rheumatology (EULAR), 2007.
13. **Arish A. Qazi**, Erik B. Dam, Paola Pettersen, Ole F. Olsen, Morten A. Karsdal, Jenny Folkesson, Claus Christiansen, "*Is Disease Cartilage more Homogeneous than Healthy Cartilage?*", In Proceedings of OARSI, 2006.

Patents

14. **Arish A. Qazi**, Erik B. Dam, Ole F. Olsen, Mads Nielsen, Paola Pettersen, Claus Christiansen, "*Pathology Indication Measure Related to Cartilage Structure and Automatic Quantification Thereof*", World Patent, 2007. [submitted]
15. **Arish A. Qazi**, Erik B. Dam, Ole F. Olsen, Mads Nielsen, Paola Pettersen, Claus Christiansen, "*Automatic Quantification of Pathology Indicating Measure Related to Cartilage Structure From Cartilage Scan Data*", Patent number GB 0618417.0 and GB 0702777.4, 2006.

Introduction

Imaging has evolved to become an essential component of clinical research and practice. The purpose of medical imaging is to provide clinicians, such as radiologists, the ability to non-invasively see the underlying anatomy and physiology of (or parts of) the human body with the aim of facilitating clinical decision making. During the past century the field of medical imaging has advanced from the conventional 2D X-Rays to a variety of complex 3D imaging techniques such as, PET, MRI, CT, and microscopy. The imaging technology market worth at \$7.7 billion in the United States, in 2007, is expected to rise to \$11.4 billion in 2011 (BCC Research). With the increasing aging population in the western world the figures will continue to rise.

The goal of medical image processing and analysis is to develop techniques that assist the clinician in both qualitative and quantitative interpretation of the useful information present in the image. The upsurge in development of cost effective imaging machinery has led to immense amount of data generation. Thus, the demand for automated methods for analysis and processing of the images is increasing; new and better techniques for quantitative analysis of the images are emerging.

This thesis presents methods for both qualitative and quantitative analysis of medical images, with a focus on magnetic resonance (MR) images. The thesis is divided in two parts; the first part presents methods that aim to analyse the articular cartilage from MR images of the knee. The degeneration of the cartilage is a hallmark process in Osteoarthritis (OA), which is a joint disease. The second part of the thesis presents methods that aim to reconstruct and analyse using diffusion tensor imaging: (1) the complex orientation of major fiber pathways in the white-matter of the brain; and (2) the orientation of collagen fibers in the articular cartilage.

This thesis is based on research papers which have either been accepted or are currently in review. Only a subset of papers are selected for which the author of the thesis is the primary author. Each part of the thesis starts with an introduction to the contents along with the background information, which is followed by the relevant research papers, each assembled as a separate chapter with its own introduction.

1.1 Purpose

OA is a disease characterized by degeneration of the articular cartilage. There is no well documented cure for OA and current treatments are directed towards relief of symptoms. One of the primary reasons is that the disease is diagnosed relatively late in its course. This is due to the lack of a non-invasive, accurate, and reproducible technique for measuring progression of OA. The purpose of the project, described in first part of the thesis, was to devise new imaging based markers, from already segmented articular cartilage, that are able to monitor progression of OA and offer a reliable detection of early OA. Moving a step forward, the next phase of the project was development of methods that aim at finding the most pathological regions in the articular cartilage.

Diffusion MRI is a recent technique that can measure the random thermal motion of water molecules in a tissue. Diffusion Tensor (reconstructed from diffusion-weighted MR images) Imaging is a widely used technique to analyse the structure of white matter fiber bundles in the brain. A fundamental drawback of the traditional diffusion tensor model is that fiber crossings in the brain cannot be modelled, and which results in erroneous fiber tracts. Therefore, the purpose was to develop methods that could robustly reconstruct white matter fiber pathways in regions of crossing fibers.

1.2 Contributions and Outline

One of the main contributions of the thesis is proposing an imaging marker, "Cartilage Homogeneity", quantified from the cartilage texture, as visualized by MRI. Cartilage homogeneity relates to early biochemical and structural changes undergoing in the cartilage and thereby is able to reliably detect early OA. For this purpose, an initial investigation of various textural measures, quantified from the cartilage, was also carried out. Results evaluating cartilage homogeneity as a possible imaging marker for OA are presented in Chapters 3 - 5.

A novel and generic framework for finding the regions of most pathological changes in a biological object is presented. Preliminary experiments have been carried out on the articular cartilage. Initial results are discussed in Chapters 6 - 7, and the framework is presented in Chapter 8.

A new multi-tensor tractography method for reconstruction of white matter fiber pathways is proposed and discussed in Chapter 10. The method is able to reconstruct fibers crossings in the corticospinal tract region of the brain. The method has been implemented to be an integral feature of the open source toolkit, TEEM (Kindlmann, 2007).

Preliminary results on the investigation of the articular cartilage structure by high-resolution diffusion tensor imaging are presented in Chapter 11.

The clinical data from the knee consists of various MR imaging modalities. The fused information from the various modalities combined with the developed methods may reveal new findings about the pathology of the disease. A first step in this direction is alignment of the data to one coordinate system. For this purpose an Affine/Rigid registration toolkit was developed. It should be noted here that due to a change in the direction of the Ph.D. research the fusion project was left as future work, but nonetheless the developed toolkit is available for online download (Qazi, 2007) and is currently being utilized by various researchers.

Part I

From Medical Images to Localization of Anatomical Changes

Osteoarthritis: Imaging and Biomarkers

Part I of the thesis consists of research papers that focus on two different, but interdependent avenues. The first three Chapters (3 - 5) are dedicated to investigating cartilage texture, from MRI, as a marker for evaluating progression of OA. Chapters 6 - 8 discuss methods that aim to find the most pathological regions in the articular cartilage, which lead to a maximum discrimination between two groups of OA subjects.

This Chapter serves as an introduction to the background information, such as: quantification, imaging, and biomarkers of OA.

2.1 Osteoarthritis (OA)

OA is a chronic disease of the joints; in fact it is the most common joint disorder. It is the second leading cause of disability in the United States, affecting the lives of more than 20 million Americans (Lawrence et al., 2008), with an estimated economic cost of \$65 billion annually (Burstein et al., 2000). Currently, there is no known cure for OA and the treatments are directed towards relief of symptoms. The prevalence of OA increases with age, and thus with an increase in longevity the impact of the disease is also expected to go up. Potential risk factors of OA have been identified to be systemic factors, such as increasing age, gender (high incidence of OA in women), and genetics, or biomechanical factors, such as obesity, occupational activity, and joint injury (Felson et al., 2000).

Although the pathology of OA is well documented the etiology of the disease is a controversial subject. OA is generally seen as a disease characterized by degeneration of the articular cartilage or abnormality of the subchondral bone

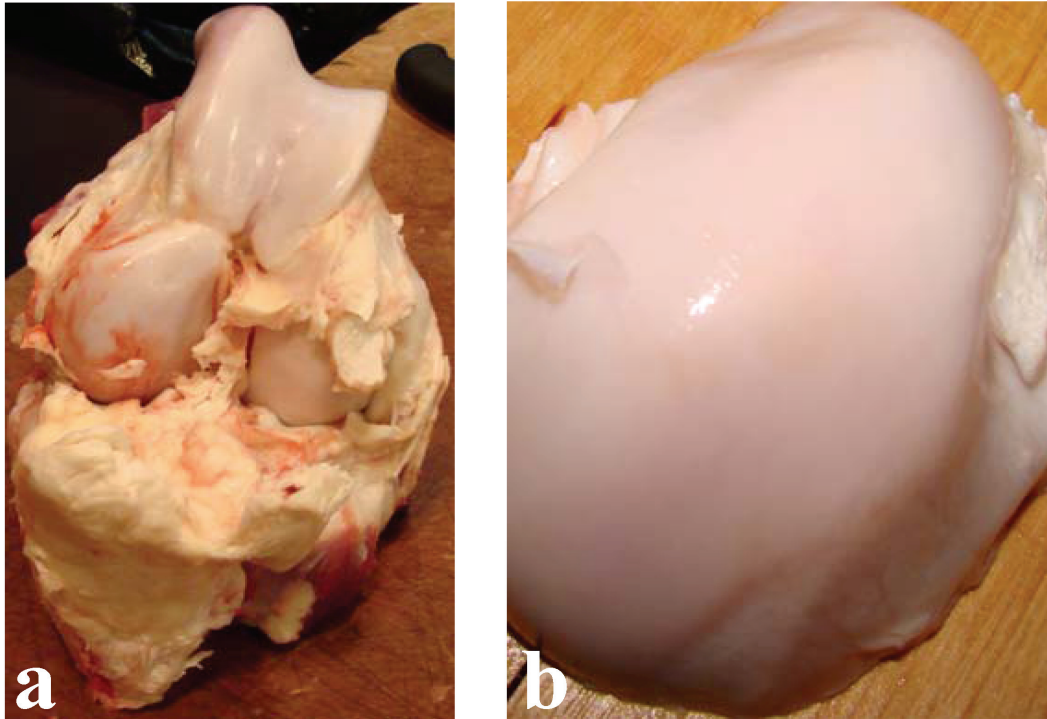


Figure 2.1: Bovine knee sample acquired from a local abattoir. (a) The knee joint: exposing the femoral and patellar cartilages (b) Femoral cartilage. We can see that the cartilage surface has a glass-like appearance, providing a smooth, elastic, friction-bearing surface that can bear enormous loads.

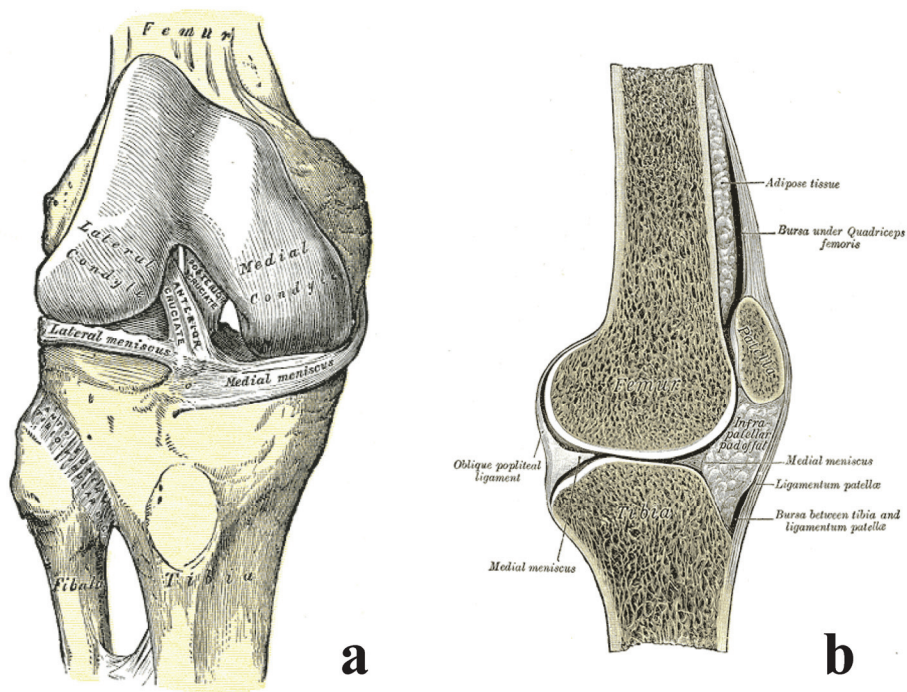


Figure 2.2: The right human knee joint. (a) Shown from the front. (b) Sagittal view. (Reproduced from, *Anatomy of the Human Body*, by Henry Gray, 20th ed; Fig. 347. and Fig. 350).

(Creamer and Hochberg, 1997). The methods presented in this thesis focus on the articular cartilage only.

Articular cartilage is a hyaline cartilage (60 - 80% of the total weight is water), with its nearly frictionless surface it is capable of resisting enormous loads and is able to distribute biomechanical forces across the bone surfaces. Figure 2.1 shows an image of a bovine knee cartilage. Cartilage is an avascular and aneural tissue that is composed predominantly of an extracellular matrix. It is generally believed that the disruption of the extracellular matrix is the process that triggers OA (McCauley and Disler, 1998). Refer to Chapter 11 for more details on the composition and layout of the cartilage matrix.

OA can occur in any joint, however, it is commonly found in weight-bearing joints, such as the knee and the hip. In this thesis the focus is on knee OA.

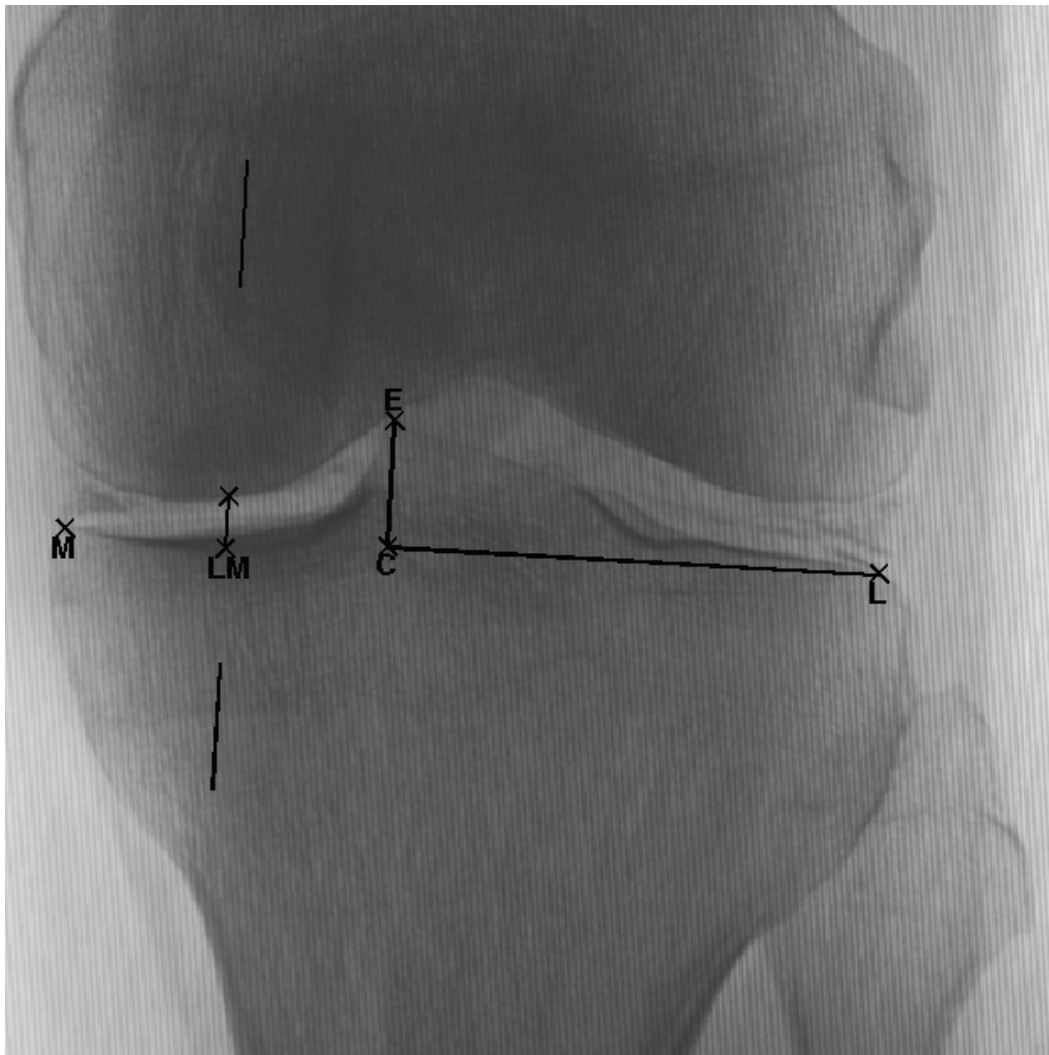


Figure 2.3: A plain radiograph with manually outlined landmarks for measuring the JSW.

Table 2.1: Kellgren and Lawrence Grading Scores

Level	Remark	Definition
0	Healthy	Healthy joint.
1	Doubtful	Doubtful JSN and minute osteophytes.
2	Minimal	Definite osteophytes and JSN.
3	Moderate	Moderate diminution of joint space and some sclerosis.
4	Severe	Joint space greatly impaired, severe sclerosis, and large osteophytes.

2.2 Knee Osteoarthritis

The knee joint, shown in Figure 2.2, is the most common site affected by OA. The current accepted gold standard for monitoring knee cartilage degeneration is Arthroscopy (Ayrar et al., 2005). Arthroscopy, however, is invasive and very expensive. Additionally, it does not allow visualization of internal cartilage structure and therefore is only limited to monitoring surface abnormalities.

The most commonly used non-invasive imaging technique for monitoring progression of OA has been radiographs; however, recently MRI has emerged as a more promising technique. Both modalities are discussed in the following section.

2.2.1 Imaging

2.2.1.1 Radiography

The most widely used imaging technique for assessment of OA is radiographs (Boegård and Jonsson, 1999). From radiographs, the *joint space width* (JSW), or the distance between the tibia and the femur bone is measured (the term *joint space narrowing* (JSN) is used to refer the longitudinal analysis of JSW). A plain radiograph is shown in Figure 2.3.

The most widely used grading scale for OA is from radiographs and is referred to as the Kellgren and Lawrence (KL) index (Kellgren and Lawrence, 1957). The scale varies from 0 – 4, where 0 is healthy and 4 is severe OA, and is determined by a radiologist who takes the appearances of JSN, osteophytes, sclerosis, and cysts into account. The grading scheme is listed in Table 2.1.

An inherent problem with using radiographs is that they can only provide information about bone changes and not cartilage deformities, since the cartilage is not visible on radiographs. Therefore, JSN is an indirect evaluation of cartilage loss.

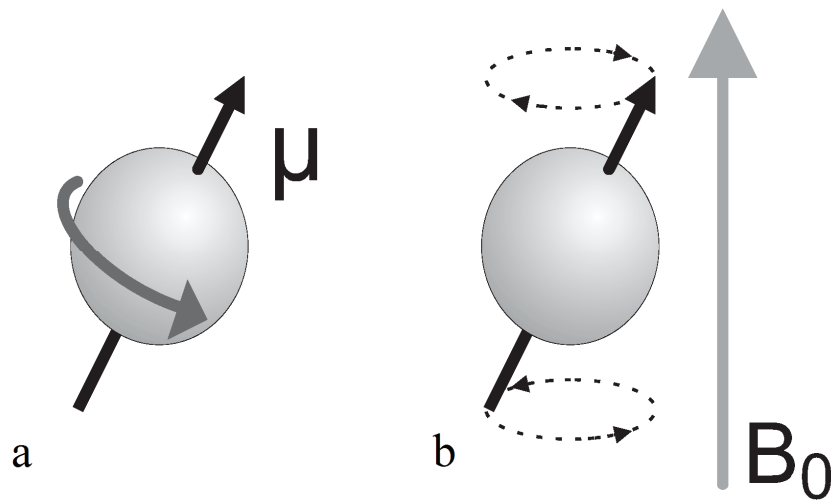


Figure 2.4: (a) A charged nuclei with a spin causes a magnetic moment μ . (b) When placed in an external magnetic field B_0 , the spins precess around B_0 with a frequency ω .

2.2.1.2 Magnetic Resonance Imaging

In recent years, Magnetic Resonance Imaging (MRI) has emerged as a popular technique for imaging of the articular cartilage (Gold et al., 2006). Primarily, this is attributed to the superior soft tissue contrast provided by MRI, apart from no radiation or harmful side effects. Moreover, the technique is non-invasive and allows direct high resolution visualizations of all the components of the joint, specifically the articular cartilage. The three-dimensional information has led to development of methods that quantify morphological information from knee MRI, such as cartilage volume and cartilage thickness, for a review see (Eckstein et al., 2006). The following section will serve as a brief review of basic MRI principles, for more details refer to (Liang and Lauterbur, 1999).

Basic Principles of MRI: A Brief Review

MRI is based on the principles of Nuclear Magnetic Resonance (NMR) (Lauterbur, 1973). It measures spatial variations in the phase and frequency of radio frequency energy that is emitted and absorbed by an object.

Protons and neutrons, in the nucleus of an atom, have an intrinsic angular momentum, called spin. A characteristic of the nuclei with an odd mass number is that they have spins in multiples of $\pm 1/2$. The charged nuclei with the spinning motion cause a magnetic dipole moment μ in the direction of the spinning axis. The strength of the moment depends on the type of nucleus. There is an abundance of magnetic nuclei in the human body (H, C, Na, P, K). The most interesting nuclei for MRI is of hydrogen, since most of

the body is constituent of water (or H_2O), and hydrogen nuclei possess a strong magnetic moment.

When placed in an external uniform magnetic field B_0 the magnetic moments tend to align with the magnetic field. The spins tend to align parallel or anti-parallel to B_0 , however, majority of the spins will be in the parallel state, or lower energy state, implying a net magnetization vector M along B_0 . The spins are not exactly aligned with B_0 and due to this misalignment the nuclei experience a torque and precess around B_0 with frequency ω . The motion is similar to that of a spinning top, precessing or wobbling in the Earth's gravitational field, as illustrated in Figure 2.4. The frequency ω is also known as the Larmor frequency and is proportional to the strength of magnetic field B_0 given as

$$\omega = \gamma B_0$$

where γ is the constant of proportionality and is known as the gyromagnetic ratio.

Next, a radio-frequency (RF) pulse B_1 , at the characteristic Larmor frequency, is applied perpendicular to B_0 , in the x-y traverse plane. The RF pulse excites the spins in the lower-energy state and causes a disruption in the equilibrium, thus there is an increase of nuclei in the anti-parallel state. This leads to a decrease in the longitudinal magnetization. This phenomenon is also known as magnetic resonance.

Due to the application of B_1 the x-y components of the net magnetization vector M are made coherent and hence cause M to tilt away from B_0 into the x-y plane. Once the RF signal is removed, the nuclei realign themselves with B_0 and thus, return to the equilibrium state. This process is known as relaxation. During relaxation the nuclei loose energy by emitting an RF signal, also known as the *free-induction decay (FID)* response signal. The FID signal is measured by a conductive coil place around the object being imaged. This measurement is processed and reconstructed to form a 3D MR image.

The relaxation process is controlled by two important parameters, the relaxation times, T_1 and T_2 . T_1 relaxation time is the time taken for magnetic moments of the displaced nuclei to return to equilibrium (realign with B_0), leading to a gradual increase in longitudinal magnetization. This is also referred to as spin-lattice relaxation. T_2 , also known as the spin-spin relaxation, measures the time required for the x-y component of the magnetization to decay leading to a decrease in traverse magnetization.

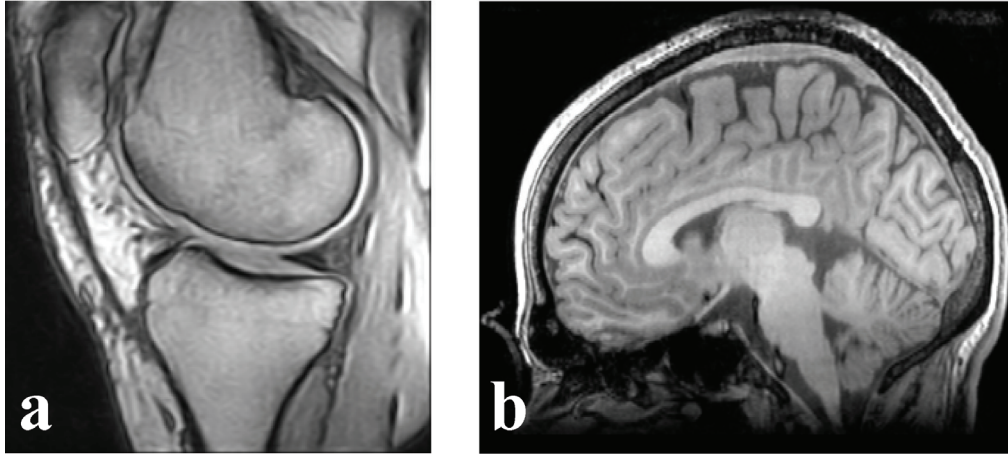


Figure 2.5: Sagittal slice from MR images of the (a) Knee, T1-weighted. (b) Brain, T2-weighted.

During the image acquisition the RF pulse B_1 is repeated at a pre-determined rate. The amount of time that exists between successive excitation pulses is known as the repetition time TR. The FID response signal can be measured at different times between the TR intervals. The time interval between the RF pulse and measurement of the FID signal is known as the echo time TE.

The contrast in MR images is controlled by adjusting TR and TE for different tissue types (Rinck, 2001). This feature of MR makes it a very powerful imaging technique. Short TR and TE results in T1-weighted images while long TR and TE leads to T2-weighted images. Figure 2.5 shows a slice acquired from MR images of the knee and the brain.

2.2.1.3 MR Appearance of Healthy Articular Cartilage

The appearance of the articular cartilage depends on the type of MR sequence. Cartilage has high signal intensity relative to joint fluid and subchondral bone on T1-weighted images, while on T2-weighted images, cartilage has a low signal intensity when compared to the joint fluid and subchondral bone (Loeuille et al., 1998).

With low-spatial resolution imaging the cartilage appears as a single homogenous layer. On high resolution images, acquired using high-field scanners, however, a laminar pattern of signal intensity can be seen (Jeffrey and Watt, 2003; Loeuille et al., 1998; McCauley and Disler, 1998). It was previously thought that this appearance was due to the different concentrations of water and proteoglycans throughout the articular cartilage, however, more recently this appearance has been attributed to the alignment of the collagen fibers in the cartilage (Burstein et al., 2000). It is a general consensus that this difference in contrast is because of the variations in T_2 across the cartilage, and which arises because of the 'magic angle' effect

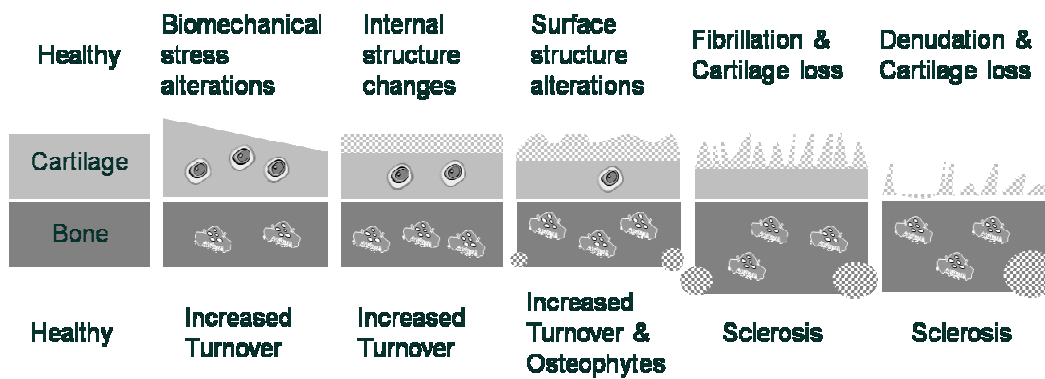


Figure 2.6: Model of OA progression in the cartilage and the bone. The figure highlights the different stages of the disease. Reproduced with permission from (Qvist et al., 2008).

(Erickson et al., 1993). This effect is due to the anisotropic orientation of the collagen fibers, which lead to an increased T_2 when the fibers are at a 55° angle with respect the main magnetic field.

2.2.2 Biomarkers

OA is a slowly progressed, chronic disease with no known cure. While the etiology of OA is still controversial it is a general consensus that OA is characterized by progressive destruction of the cartilage tissue and the subchondral bone. Research is on-going to discover drugs that could slow down this process. These drugs are referred to as disease modifying osteoarthritis drugs (DMOADs) (Abadie et al., 2004). However, to date no DMOADS have been discovered that are able to reverse the pathogenesis of OA.

Biomarkers are used to indicate or measure a particular state of a biological process. To test the efficacy of a certain drug biomarkers that are sensitive to the outcome of the disease are needed. Current diagnosis of OA relies on clinical history and radiography. These techniques, however, rely on cartilage loss. Degeneration of the articular cartilage is an irreversible process because the cartilage has limited capacity to regenerate or repair itself (McCauley and Disler, 1998). Therefore, using radiographic outcome variables for clinical trials of potential DMOADS might not be feasible because they are insensitive to detection of early biological changes undergoing in the cartilage. By the time the first changes in radiological OA are detected, on average 13% of the cartilage has already degenerated (Jones et al., 2004). Additionally, radiographic changes are slow and typically last 2 - 3 years (Jones et al., 2004), resulting in expensive clinical trials.

Therefore, sensitive clinical diagnostic biomarkers of cartilage degradation are needed. Markers that are able to target early OA would be able to identify

individuals at high risk and predict response to early interventions. Biochemical markers such as CTX-I (bone turnover) and CTX-II (cartilage turnover) have been developed (Schaller et al., 2005), however, in this thesis we will only focus on imaging markers, quantified from MRI.

Recently, MRI is widely being used to image the knee joint, and quantitative morphological measurements from MRI are being evaluated as a surrogate end point. OA is a complex, evolutionary disease and the pathogenesis of OA may be represented by the model proposed in (Qvist et al., 2008), shown in Figure 2.5. Current imaging markers from MRI such as Cartilage Volume and Cartilage Thickness focus on cartilage loss, which is directed to the late stages of the disease, as shown in Figure 2.5. However, as MRI research unfolds new parameters are emerging that aim to quantify early biochemical and physiological changes undergoing in the articular cartilage. Some of these include quantitative assessment of water and collagen concentration by MRI T_2 relaxation time (Lüssea et al., 2000), concentration of glycosaminoglycan (GAG) molecules by delayed Gadolinium Enhanced MRI of the Cartilage (dGEMRIC) (Tiderius et al., 2003), analysis of cartilage structure by diffusion tensor imaging (DTI) (Filidoro et al., 2005), measurement of proteoglycan content by sodium MRI (Reddy et al., 1998). For more details, refer to excellent reviews by (Burstein, 2006; Burstein et al., 2000; Gold et al., 2006).

In this thesis, the focus is to develop reliable imaging based markers of OA from MRI that focus on early detection and progression of the disease. Chapters 3 - 4 are dedicated to investigating cartilage texture, quantified from existing segmentations of the articular cartilages from MRI, as a possible marker for early detection and progression of OA.

When recruiting in clinical trials it is essential to be able to differentiate progressors of the disease from non-progressors. Currently both groups are included and this may result in loss of sensitivity to the measurement of disease progression. Thus, Chapter 5 looks at cartilage homogeneity as a possible marker for evaluating the progression in OA. Chapters 6 - 8 presents methods that based on an imaging marker, aim to find the most affected regions of the articular cartilage during the pathogenesis of OA.

Automatic Quantification of Cartilage Homogeneity

Arish A. Qazi, Ole F. Olsen, Erik B. Dam, Jenny Folkesson, Paola Pettersen, and Claus Christiansen. In Proceedings of MICCAI, 2006.

Abstract

Osteoarthritis (OA) is a degenerative joint disease that involves the wearing down of the articular cartilage. Typical setbacks have been quantifying progression and early detection of the disease. In this study, we develop a fully automatic method for investigating cartilage homogeneity as a potential maker for disease quantification. The method is evaluated on 114 manually and automatically segmented knee MR images from subjects with no, mild, or severe OA symptoms. To measure homogeneity we characterize the tibial and femoral compartments in each cartilage sheet by several statistical measures, and then evaluate their ability to quantify OA progression. This is done by statistically testing the discriminatory power of each measure in separating the groups of healthy subjects from those having OA. Our chosen measure outperforms a popular measure like cartilage volume in statistically discriminating the two groups. Through a scan-rescan evaluation, we show that the measure is reproducible.

3.1 Introduction

The majority of the elderly population will at some point encounter osteoarthritis (OA) resulting in pain and reduced range of motion in mainly the knee and hip joints. For the worst cases, joint replacement surgery is even required. Currently, available treatments are directed towards relief of symptoms and at present no drugs have shown to consistently modify joint structure or reverse joint pathology (Altman, 2004). New, accurate and precise methods are needed in order to quantify the disease progression in clinical studies that determines the effect of potential treatments. A hallmark process in OA is cartilage breakdown. Therefore, typical disease progression quantifications have traditionally been:

- Joint gap (between the femur and tibia on the knee joint) is an indirect measure of cartilage thinning from radiographs (X-ray) where the cartilage itself is not visible (Brandt et al., 1991).
- Cartilage volume and thickness measured from magnetic resonance imaging (MRI) where the cartilage is visible (Folkesson et al., 2005b; Graichen et al., 2004).

However, before thinning even begins, the cartilage loses its firm structure (Hollander et al., 1995). This structure is composed by the three layers of collagen fibers that allows absorption/expulsion of water and thereby ensures shock absorption. When this structure is breaking down, the first stage is swelling of the cartilage which is followed by thinning in the later stages (Calvo et al., 2004). Therefore, measures focusing on volume or thickness may not be adequate to capture the early stages of OA.

The early loss of integrity could ideally be measured directly by quantifying the decrease in alignment of the collagen fibers in the three layers of the cartilage. In the future, this will possibly be done through analysis of very high resolution diffusion tensor MRI (Filidoro et al., 2005). Due to limitations in resolution and acquisition times, we settle for analysis of the intensities observed in regular MRI as a first step. Instead of measuring integrity of cartilage layer alignment directly, we therefore quantify “cartilage homogeneity”.

In this paper, we investigate a number of measures for measuring cartilage homogeneity and evaluate their ability to quantify OA progression. Since there is no ground truth for cartilage homogeneity available, we cannot evaluate accuracy directly. Possibly, a validation could be performed based on histological slices of cartilage. However, due to the highly invasive nature of this, it is not suitable as a standard in vivo evaluation technique.

We therefore evaluate accuracy indirectly by investigating the ability of the measures to separate healthy knees from knees with some degree of OA from

a collection of 114 knees. We evaluate the robustness of the method by comparing the measures obtained from both manual and automatic segmentations of the knees. Finally, we evaluate the reproducibility of the method through a scan-rescan evaluation where measures obtained from 31 pairs of scans acquired a week in-between are compared. The investigation shows that a subset of the proposed measures separates healthy from OA more accurately than volume. Furthermore they are able to detect early stages of OA.

3.1.1 Related Work

Several studies have shown correlation between OA progression and the mean intensities from T2 MR scans (Naish et al., 2004). Analysis of structure has previously been applied to anatomical structures other than the cartilage. For example, the structure of trabecular bone can be quantified by fractal signature analysis (Buckland-Wright et al., 1996). Texture analysis is also used for osteoporosis progression analysis in the form of Minkowski functionals and scaling vectors (Raeth et al., 2006).

The thin cartilage structure with a typical thickness of only a few voxels complicates the analysis. However, a few studies have done early feasibility studies on the use of diffusion tensor MRI (Basser et al., 1994b) for cartilage analysis. In a small in vitro study, it has been shown that the eigenvector orientations of the local diffusion tensor can separate cartilage layers that to some degree reflects the alignment of the collagenous fiber network in the articular cartilage (Filidoro et al., 2005).

3.2 Methods

3.2.1 Imaging

A total of 114 right and left knees were examined by radiography and MRI. Using radiographs (X-rays) these 114 knees were classified by a radiologist as a 0 - 4 on the Kellgren-Lawrence (KL) (Kellgren and Lawrence, 1957) index where KL 0 represents healthy and KL 4 severe OA. Among the 114 subjects, 51 are healthy (KL = 0) and 63 have OA (KL > 0). MR Image acquisition was done on an Esaote C-Scan low field 0.18T clinical scanner. The imaging sequence consisted of 3D, T1-weighted Gradient-Echo acquisition (GRE) (flip angle = 40° , $T_R = 50$ ms, $T_E = 16$ ms). The field of view (FOV) was 120 mm. The scans were made through the sagittal plane with the image matrix of 256 x 256 pixels, yielding a pixel size of 0.49 mm. A 110 contiguous slices, 0.8mm thick (0.7 - 0.9mm) were acquired in approximately 10 minutes.

The medial compartment of the tibial and femoral cartilage sheets was manually segmented by a radiologist. For evaluating the robustness of our method the same set of cartilage sheets was automatically segmented using the method described in (Folkesson et al., 2005a). The segmentation is based

on a three class approximate k-Nearest Neighbor (kNN) classification scheme. For reproducibility evaluation the MR protocol was repeated a week later on a subset of 31 knees

3.2.2 Cartilage Homogeneity

We define cartilage homogeneity to be a measure of the signal intensity variation inside the segmented cartilage compartment. To quantify homogeneity we choose two types of statistical measures.

3.2.2.1 First Order Statistical Measures

The first order statistical measures rely on approximating the probability of observing a particular intensity at a randomly chosen location in the image. The measures are calculated from the gray scale histogram of the image defined by $H(i) = n_i$; where $n_i (i = 0, 1, \dots, L-1)$ represents the number of occurrences for signal intensity i and L represents the number of distinct gray levels in the cartilage. Using the above formulation of the histogram we define the following measures:

Mean (average intensity): $\mu = \sum_{i=0}^{L-1} iH(i)$

Standard deviation (contrast of image): $\sigma = \sqrt{\sum_{i=0}^{L-1} (i - \mu)^2 H(i)}$

Uniformity (energy of image): $\mu = \sum_{i=0}^{L-1} H(i)^2$

Normalized Entropy (randomness): $-\sum_{i=0}^{L-1} \frac{H(i) \log H(i)}{\log(q)}$

When measuring entropy and uniformity the histogram is smoothed by decreasing the number of bins. $H(i)$ then corresponds to the number of occurrences of intensities within a range defined by the bin width. The bin width is determined by the Freedman-Diaconis rule (Freedman and Diaconis, 1981) and is chosen to be 100. Furthermore, to ensure invariance to change in the image quantization levels we normalize entropy by logarithm (base 2) of the quantization levels ($q = 4096$).

3.2.2.2 Second Order Statistical Measures

The second order statistical measures rely on the joint probability distribution of pairs of voxels in the image. This involves the estimation of the discrete second-order probability function $C_{d\theta}(i, j)$ which represents the probability of occurrence of a voxel pair with gray levels i and j given the spacing between

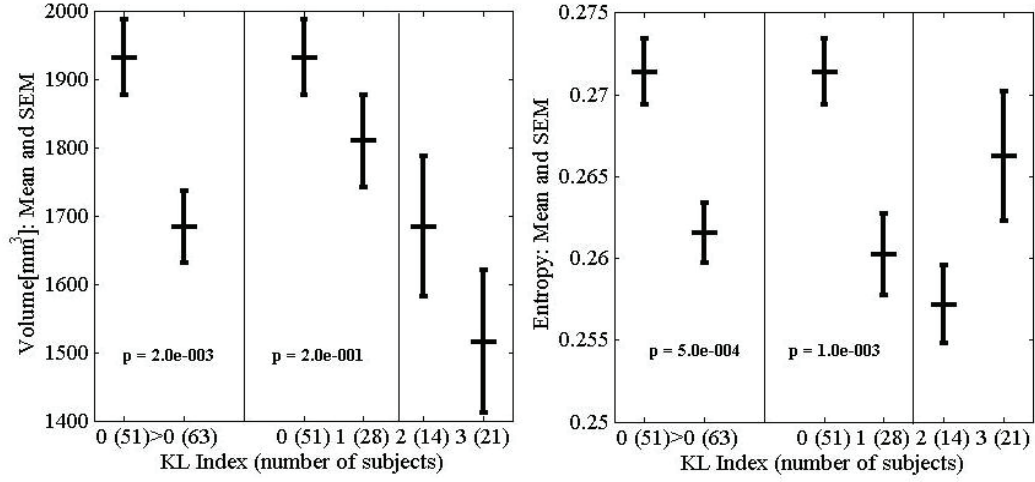


Figure 3.1: Comparison of volume (left) and entropy (right) measure as a function of the KL index.

the pair of voxels is d along a given direction θ . $C_{d\theta}(i, j)$ is called the gray level co-occurrence matrix (GLCM) (Haralick et al., 1973). Due to the size of the knee cartilage (few mm thick) we only consider immediate neighboring voxels. Furthermore, in order to reduce the dimensionality we assume that the joint probabilities are directional independent. Therefore, we average the $L \times L$ GLCM, where L is the number of gray levels in the image.

In order to quantify the spatial dependence, (Haralick et al., 1973) suggests 14 measures, each of which is extracted from the GLCM. A subset of the four most relevant measures are chosen:

$$\text{Contrast: } \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i-j)^2 C[i, j], \text{ Homogeneity: } \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{C[i, j]}{1+|i-j|}$$

$$\text{Correlation: } \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{(i-\mu)(j-\mu)C[i, j]}{\sigma^2}, \text{ Energy: } \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} C[i, j]^2$$

3.3 Results

3.3.1 Manual Segmentations

The measures for both tibial and femoral cartilage sheets for each knee are grouped according to the KL indexes, healthy ($KL = 0$) or OA ($KL > 0$). To determine the significance of each the measure we use the t-test, the null hypothesis is that the two groups (healthy and OA) have the same mean. The null hypothesis can be rejected if the resulting p-value is less than a chosen level of significance α . Using $\alpha = 0.05$, the following measures succeeded to

Table 3.1: P-values for separating KL 0 from KL > 0
(manual segmentations of the tibial cartilage)

Measure	p-value
Volume	0.002
Mean signal intensity	0.1
Standard deviation (SD)	0.0006
Uniformity	0.002
Entropy	0.0005
GLCM: Contrast	0.007
GLCM: Correlation	0.8
GLCM: Energy	0.3
GLCM: Homogeneity	0.1

discriminate healthy from OA: from first order measures it was standard deviation, uniformity and entropy, and from second order it was contrast.

Table 3.1 lists the p-values for the first and second order measures for the tibial cartilage. The p-values are low suggesting that the measures can significantly discriminate healthy from OA. Volume of the knee cartilage has recently emerged as a popular measure for monitoring OA progression from MRI (Raynauld et al., 2006). To assess and compare the discriminatory power of our measures we also include p-values for volume, as in (Folkesson et al.,

Table 3.2: P-values for separating KL 0 from KL > 0
(manual segmentations of the femoral cartilage)

Measure	Volume	SD	Uniformity	Entropy	GLCM: Contrast
Full Femoral cartilage	0.3	0.4	0.06	0.2	0.1
Load bearing region	0.3	0.05	0.01	0.02	0.1

2005b) volume of each cartilage is normalized by width of the subject's tibial plateau. Figure 3.1 shows the comparison between volume and entropy in capturing OA progression (due to space constraints we only show the comparison between volume and entropy). We can see that entropy ($p = 0.0005$) outperforms volume ($p = 0.002$) in separating healthy from OA. Furthermore, it can also detect early OA. That is by separating KL 0 from KL 1 ($p = 0.001$).

The first row of Table 3.2 lists the results for the femoral cartilage. Due to space constraint the results only list a subset of the most significant measures. The p-values are high which means that the measures are not able to discriminate the two groups. However, it was recently shown that accuracy of the thickness measurements from MR images is better in the weight bearing regions of femur (Koo et al., 2005). This is because the weight bearing region of the femur sustains contact with the tibia during the gait cycle. Thus, we approximate the load bearing region by the intersection of the two cartilage sheets. This is done by iteratively applying a 3D morphological dilation operator to the tibial cartilage region. The second row of Table 3.2 lists the results for the load bearing region of the femoral cartilage. The results slightly

Table 3.3: P-values for separating KL 0 from KL > 0
(automatic segmentations of the tibial and femoral cartilages (load bearing))

Measure	Volume	SD	Uniformity	Entropy	GLCM: Contrast
Tibial cartilage	0.001	0.00004	0.0002	0.00004	0.0003
Femoral cartilage	0.004	0.02	0.003	0.002	0.1

improve. Subsequent results, for the femoral cartilage, are estimated only for the load bearing region.

3.3.2 Automatic Segmentations

The first and second row of Table 3.3 shows the results for automatically segmented tibial and femoral (load bearing) cartilage sheets. The p-values are low suggesting that the measures succeed to discriminate the two groups. For example, entropy ($p = 0.00004$) outperforms volume ($p = 0.001$) in separating healthy from OA.

3.3.3 Reproducibility

Table 3.4: Reproducibility evaluation
(percentage difference (%) and correlation coefficient (r) for manual(m) and automatic(a) segmentation)

Type	Volume	SD	Uniformity	Entropy	GLCM: Contrast
m-Tibial (%)	11.1	9.3	9.0	4.3	21.9
m-Tibial (r)	0.82	0.82	0.78	0.78	0.81
m-Femoral (%)	9.6	12.7	12.3	4.8	32.1
m-Femoral (r)	0.91	0.6	0.63	0.64	0.55
a-Tibial (%)	10.1	8.1	8.3	3.6	16.1
a-Tibial (r)	0.75	0.81	0.8	0.78	0.86
a-Femoral (%)	45.1	11.7	12.3	4.8	27.5
a-Femoral (r)	0.86	0.65	0.62	0.64	0.63

To assess the reproducibility of the measures we compute the mean percentage difference (%) and the correlation coefficient (r). Table 3.4 lists the reproducibility results for both the manual and automatic segmentations.

The correlation coefficient can be close to 1 even if there is a considerable variation in the data (Altman and Bland, 1983). An alternate way is visual assessment of the relationship of two measures; by plotting their differences against their mean. This resultant plot is known as a Bland-Altman plot (Altman and Bland, 1983). According to Bland and Altman; if 95% of the differences lie within two standard deviations then we can conclude that the method is precise. Figure 3.2 shows a Bland-Altman plot for entropy of both manually segmented tibial and femoral cartilages. The percentage difference for entropy is low (less than 5%). Therefore, we can conclude that the measure is highly reproducible.

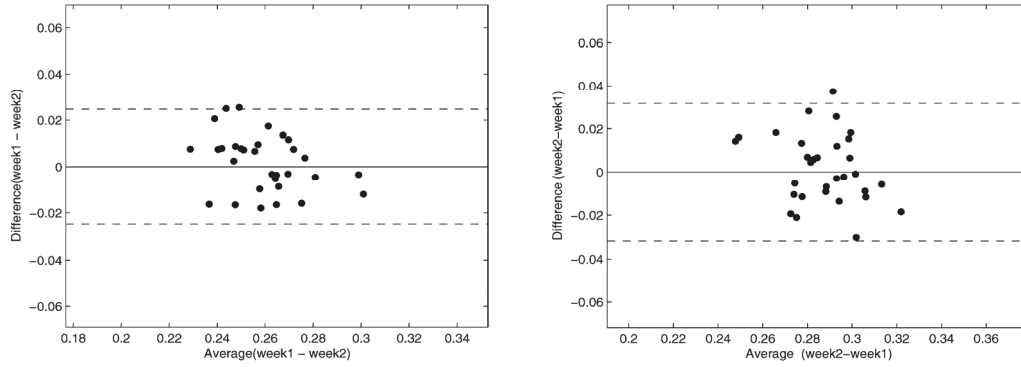


Figure 3.2: Bland-Altman plot for entropy for manual segmentations of the tibial (left) and the femoral (right) cartilage sheets. The dotted line represents two standard deviations of the differences.

3.4 Conclusion

The results of our evaluation show that cartilage texture might not be homogenous and may be utilized as a marker for disease progression. We quantify homogeneity by characterizing the manual and automatic segmentations of 114 knee cartilages by several statistical measures. To determine if the measures can quantify OA progression we do a t-test to evaluate their ability in separating groups of healthy subjects from those having OA. From a subset of measures that succeed we choose entropy to be the measure that (tibia: manual segmentation ($p = 0.0005$), automatic segmentation ($p = 0.00004$)) can discriminate healthy from OA more confidently than cartilage volume (tibia: manual segmentation ($p = 0.002$), automatic segmentation ($p = 0.001$)). Furthermore, the measure is also able to detect early progression of OA. Moreover, the measure is reproducible (tibia: % ≤ 4.3). Future work may involve:

- Probing the surroundings of the cartilage.
- Longitudinal studies.
- Comparison with histological analysis of cartilage structure.

Separation of Healthy and Early Osteoarthritis by Automatic Quantification of Cartilage Homogeneity

Arish A. Qazi, Jenny Folkesson, Paola Pettersen, Morten A. Karsdal, Claus Christiansen, and Erik B. Dam. Osteoarthritis and Cartilage, 2007.

Abstract

Cartilage loss as determined either by magnetic resonance imaging (MRI) or by joint space narrowing in x-rays is the result of cartilage erosion. However, metabolic processes within the cartilage that later result in cartilage loss, may be a more accurate assessment method for early changes.

Early biological processes of cartilage destruction are among other things, a combination of proteoglycan turnover, as a result of altered charge distributions, and local alterations in water content (edema). As water distribution is detectable by MRI, the aim of this study was to investigate cartilage homogeneity visualized by MRI related to water distribution, as a potential very early marker for early detection of knee osteoarthritis (OA).

Design: 114 right and left knees from 71 subjects aged 22 to 79 years were scanned using a Turbo 3D T₁ sequence on a 0.18T MRI Esaote scanner. The medial compartment of the tibial cartilage sheet was segmented using a fully automatic voxel classification scheme based on supervised learning. From the segmented cartilage sheet, homogeneity was quantified by measuring entropy from the distribution of signal intensities inside the compartment. For each knee an x-ray was acquired and the knees were categorized by the Kellgren and Lawrence (KL) Index and the joint space width (JSW) were measured.

The P -values for separating the groups by each of JSW, cartilage volume, cartilage mean intensity, and cartilage homogeneity were calculated using the unpaired t-test.

Results: The P -value for separating the group diagnosed as KL 0 from the group being KL 1 based on JSW, volume and mean signal intensity the values were $P = 0.9$, $P = 0.4$ and $P = 0.0009$ respectively. In contrast, the P -value for homogeneity was $P = 0.0004$. The precision of the measures assessed, as a test-retest root mean square coefficient of variation (RMS-CV%) was 3.9% for JSW, 7.4% for volume, 3.9% for mean signal intensity and 3.0% for homogeneity quantification.

Conclusion: These data demonstrate that the distribution of components of the articular matrix precedes erosion, as measured by cartilage homogeneity related to water concentration. We show that homogeneity was able to separate early OA from healthy individuals in contrast to traditional volume and JSW quantifications. These data suggest that cartilage homogeneity quantification may be able to quantify early biochemical changes in articular cartilage prior to cartilage loss and thereby provide better identification of patients for OA trials who may respond better to medicinal intervention of some treatments. In addition, this study supports the feasibility of using low-field MRI in clinical studies.

Keywords—Knee Osteoarthritis, Early Detection, Magnetic Resonance Imaging, Homogeneity.

4.1 Introduction

Osteoarthritis (OA) is a degenerative joint disease that is a major cause of disability. Degeneration of the articular cartilage in combination with an altered subchondral compartment are key features of OA (Conaghan et al., 2006; Ding et al., 2005). In terms of quality of life and chronic disability OA is second only to cardiovascular diseases (Haq et al., 2003; McCauley and Disler, 1998). It is estimated that more than one-third of the population above the age of 35 will at some point in their lives experience OA (McCauley and Disler, 1998). At present there is no cure for OA as no drugs have been consistently shown to modify joint structure or even reverse joint pathology in face of the currently available treatments that are directed towards relief of symptoms (Altman, 2004). Research is on-going to discover disease-modifying anti-osteoarthritis drugs/agents (DMOADs) (Abadie et al., 2004). However to assess the effectiveness of DMOADs we need to quantify the structural changes undergoing in the cartilage during the early stages of the disease and for that early diagnosis of OA is essential.

The current accepted standard for diagnosing knee OA and monitoring progression is measurement of the joint space width (JSW) (between the femur and tibia on the knee joint) from radiographs (Altman et al., 1996). This is an indirect evaluation as the cartilage is not visible in X-ray and is also potentially prone to diagnosing the disease relatively late in its course (Calvo et al., 2001). Recently, magnetic resonance imaging (MRI) has received much attention in assessment of the articular cartilage. This can be primarily attributed to the fact that MRI is non-invasive, provides excellent soft tissue contrast, high spatial resolution and moreover the articular cartilage can be directly visualized and quantified non-invasively from the MR scans (Eckstein et al., 2006; Pessis et al., 2003). This provides valuable information with regard to morphological and possibly biochemical parameters, that may be associated with the integrity of the articular cartilage. Quantitative measurements from MRI like cartilage volume and thickness measures are now being widely used to monitor progression of OA (Dam et al., 2005; Folkesson et al., 2005a; Pakin et al., 2002; Solloway et al., 1997).

The pathology of OA involves changes in both the subchondral bone and articular compartment. Currently, late stage detection methods such as JSW or Kellgren & Lawrence (KL) score represent subchondral bone changes as well as cartilage thinning (Buckland-Wright, 2004). Even early OA as detected by the KL score, may be the results of long ongoing biochemical processes leading to both bone and cartilage alterations. Thus, these techniques may not sufficiently identify patients which may have their cartilage disease attenuated or even reversed. Cartilage loss is only secondary to important biochemical changes in the articular cartilage. Prior to cartilage fibrillation, local edema and swelling are well-recognized features (AlHadlaq et al., 2004; Calvo et al., 2001; Calvo et al., 2004; Maiotti et al., 2000). These intrinsic changes in cartilage water content and spatial distributions are most likely

predecessors to the traditional cartilage loss detected by MRI and x-ray based techniques.

In this study, we wanted to investigate whether MRI signal intensity measures, related to intrinsic water distributions, which may be a result of mal-metabolism in the articular cartilage, would be a sensitive marker for very early detection of OA.

4.2 Materials and Methods

4.2.1 Population

The population for this study was formed from two subpopulations. The first was a normal population selected to have no or only minor OA symptoms with a broad age range from 22 to 73 years. The second was a group with known OA symptoms of various degrees. This composition was designed to ensure that a large healthy group could be compared to both a considerable group with early OA as well as a considerable group with more progressed OA. Secondly, the full population was thereby designed to be fairly representative of the general population.

A total of 76 subjects (34 men and 42 women) from 22 to 79 years old were recruited to participate in the present study. Selection criteria ensured that none of the subjects had previous knee joint replacement, inflammatory arthritis in the knee, or any contraindication for performing MRI examination. Subjects underwent both clinical and radiological examinations and classified into groups of healthy subjects or OA patients according to the ACR (American College of Rheumatology) definition of OA. From the total of 152 knees, 13 were excluded due to poor image quality in either radiographs or MRI. From these 139 knees, 25 were used to train the automatic segmentation method, so a total of 114 knees (Table 4.1) were used for evaluation of the methods in the following.

All participants signed approved information consent and the study was carried out in accordance with the principles of the Helsinki Declaration II and European Guidelines for Good Clinical Practice. The study protocol was approved by the local Ethical Committee.

In this study, we focused on the medial compartment of the tibial cartilage since previous studies have shown that OA is more often observed in this compartment (Dunn et al., 2004).

4.2.2 Imaging Protocol

A total of 152 right and left knees were examined by radiography and MRI. Using radiographs (X-rays) these 152 knees were classified by a radiologist as 0 - 4 on the Kellgren-Lawrence (KL) (Kellgren and Lawrence, 1957) index where KL 0 represents healthy and KL 4 severe OA (It should be noted here

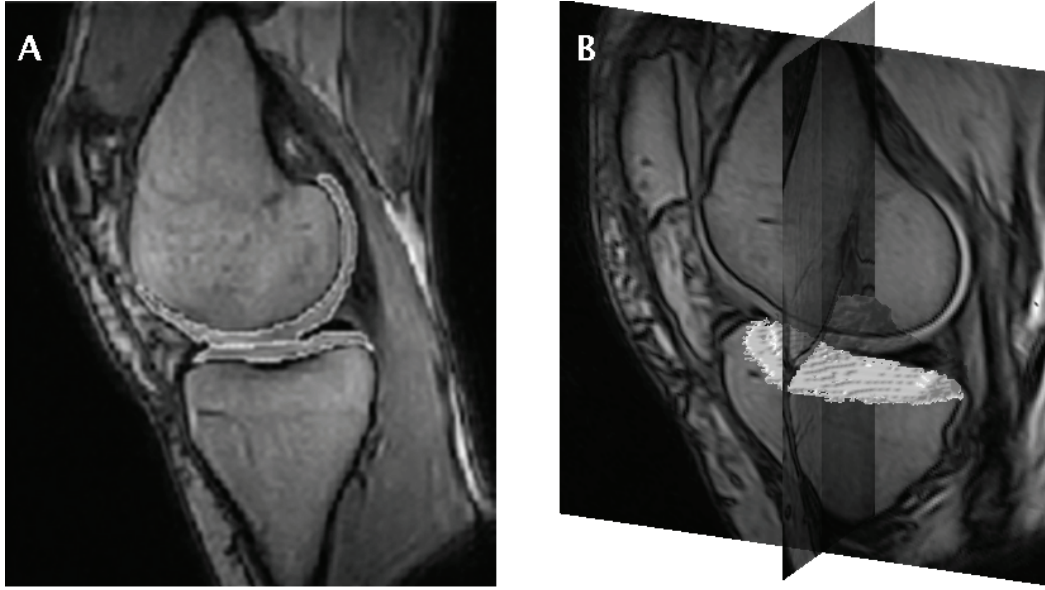


Figure 4.1: (a) Automatically segmented Sagittal slice (b) Cross section view of a segmented tibial medial cartilage sheet.

that throughout the paper “healthy” refers to being radiologically healthy i.e. being classified by a radiologist as having a KL score of 0). The radiologist also measured the JSW as the narrowest gap between the tibia and the femur in the medial compartment. The radiologist also marked the most medial and lateral points on the tibial plateau (excluding osteophytes). These points define the width of tibial plateau, which is a measure of the size of the knee.

MRI Image acquisition was done on an Esaote C-Scan low field 0.18T clinical scanner. The imaging sequence consisted of 3D, T1 weighted Gradient-Echo acquisition (GRE) (flip angle = 40° , $T_R = 50$ ms, $T_E = 16$ ms). The field of view (FOV) was 120 mm. The scans were made through the sagittal plane with the image matrix of 256×256 pixels, yielding a pixel size of 0.49 mm. A 110 contiguous slices, 0.8mm thick (0.7 - 0.9mm) were acquired in approximately 10 minutes.

For reproducibility evaluation, the protocol was repeated a week later on 31 knees. Table I lists the characteristics of these knees.

4.2.3 Automatic Cartilage Segmentation

The medial compartment of tibial cartilage sheets were automatically segmented using voxel classification based on supervised learning (Folkesson et al., 2005a). For the classification we used an approximate k NN classifier, which was implemented in an ANN (Approximate Nearest Neighbour) framework. To account for position variations of subjects in the scanner the centre of mass of each scan was shifted to the centre of mass of the scans in

Table 4.1: Characteristics of the evaluation set (knees = 114, participants = 71)*

Factor	KL 0	KL 1	KL 2	KL 3
Number of knees (<i>K</i>)	51	28	14	21
Age (yrs)	45.2 ± 15.7	58.7 ± 11.2	67.5 ± 7	66.3 ± 6.7
Females (%)	60%	71%	43%	48%
Weight (kg)	72.4 ± 12.8	71.6 ± 10.1	89.8 ± 16.6	81.1 ± 11.6
Height (m)	1.7 ± 0.09	1.7 ± 0.08	1.7 ± 0.09	1.6 ± 0.07
BMI (kg/m ²)	24.3 ± 3.6	24.8 ± 3.2	30.3 ± 4.2	28.8 ± 3.5
Reproducibility 2nd Visit (knees = 31)				
Number of knees (<i>K</i>)	11	13	2	5
Age (yrs)	52 ± 18.9	66.1 ± 7.2	65.2 ± 0.0	68 ± 4.5
Females (%)	55%	77%	0%	0%
Weight (kg)	69.8 ± 17.1	69.3 ± 8.0	87 ± 0.0	84 ± 11.6
Height (m)	1.7 ± 0.1	1.68 ± 0.04	1.68 ± 0.0	1.75 ± 0.05
BMI (kg/m ²)	23.9 ± 4.7	24.5 ± 2.2	30.9 ± 0.0	27.3 ± 3.0

* Values are mean ± SD. KL refers to the Kellgren and Lawrence Index

the training set (Folkesson, 2006). Furthermore the computational time was significantly reduced by incorporating an efficient classification scheme (Dam et al., 2006a). Figure 4.1 shows an MR scan slice illustrating the automatic segmentation. From the segmented cartilage sheet cartilage mean signal intensity and volume for each subject was computed. The volume measure was normalized by the width of the subject's tibial plateau (width of the bone). Additionally standard deviation of the signal intensities was also computed since standard deviation is another way of measuring dispersion or randomness.

4.2.4 Quantification of Homogeneity

Homogeneity was quantified by measuring entropy. Entropy is well-known in information theory (Shannon, 1948) as a measure of information content (or inversely randomness) present in the data. Entropy is measured from the signal intensity histogram, which represents the distribution of intensities present inside the cartilage compartment.

A histogram H can be defined by the following formulation: $H(i) = n_i$; where $n_i (i = 0, 1, \dots, L-1)$ represents the number of occurrences for signal intensity i and L represents the number of distinct grey levels in the cartilage. The histogram is normalized by the total number of intensities N such that the histogram represents the probability distribution of the signal intensities – and thereby also becomes invariant to the cartilage volume.

Furthermore, the intensity range was divided into equal sized bins. The bin width was chosen to be 100 using the Freedman-Diaconis rule (Freedman and Diaconis, 1981). Using this normalized binned histogram we calculated entropy as:

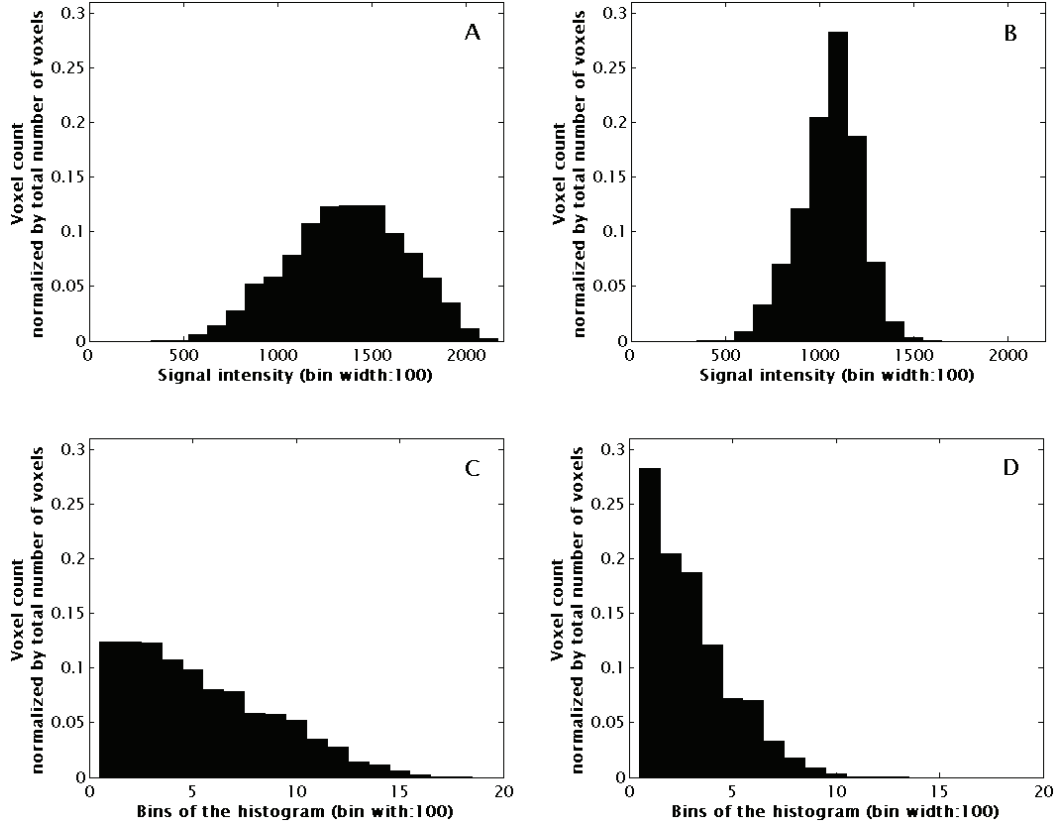


Figure 4.2: (a) Histogram of knee with highest entropy value in the dataset (KL 0) (b) Histogram of knee with lowest entropy value (KL 3) (c) The same histogram in (a) but sorted by the number of bins (d) The same histogram in (b) but sorted by the number of bins. The more diseased you are the lower will be the entropy.

$$E = - \sum_{i=0}^{N-1} H_i \log(H_i) \quad 4.1$$

Entropy quantifies cartilage with fewer, more dominant intensities as being more homogeneous. In the extreme, the most homogeneous distribution will only have a single intensity present – such a histogram can be described with very little information and therefore has minimal information content and minimal entropy value. In contrast, the higher the entropy value the more heterogeneous the cartilage will be. As an example, Figure 4.2(a) and Figure 4.2(b) shows the histograms of knee cartilage with highest entropy as well as with the lowest entropy in the dataset. When assessing the information content, the specific intensities and their ordering in the histogram are irrelevant – this is illustrated by Figure 4.2(c) and Figure 4.2(d) where the same histograms are sorted by the bin size. The entropy value depends on whether many in-frequent intensities are present (heterogeneous, high entropy) or whether relatively few frequent intensities are dominant (homogeneous, low entropy). This is clearly visible in the sorted histogram in

Figure 4.2(c) and Figure 4.2(d). Thereby, unlike a simpler measure like intensity standard deviation, entropy is not assuming that the intensity values follow a Gaussian distribution.

4.2.5 Statistical Analysis

To evaluate the significance of homogeneity as a biomarker for early detection of OA, the knees were divided in groups based on the KL index; and using the unpaired t-test, P -values for separating the groups based on entropy were calculated. A 0.05 level was used throughout the paper for statistical significance. The significance levels were compared to the significance levels of separating the same groups using JSW, traditional volume quantification and the mean signal intensity.

The overall precision of the method was assessed using the test-retest root mean square coefficient of variation (RMS-CV%) of the entropy values (Gluier et al., 1995). For each subject the CV is defined as the standard deviation of a series of experiments divided by the mean. The overall CV is expressed as an RMS of the subjects CVs.

4.3 Results

4.3.1 Cartilage Homogeneity Measure

To assess the homogeneity we measured T_1 intensities, which are correlated to distribution of water, as shown in Figure 4.2(a) and Figure 4.2(b) of a healthy knee compared to that of a KL 3 knee. From these the intensity standard deviation (SD) was seen to be lower in diseased cartilage compared to that of healthy cartilage (SD healthy versus SD OA, $P < 0.001$). For further analysis we used entropy, which quantifies homogeneity as the information content, reflecting the complexity, of the intensity histogram as illustrated in Fig. 2. The entropy was lower in diseased cartilage compared to healthy (Entropy healthy versus Entropy OA, $P < 0.001$). The precision of SD based on test-retest RMS-CV% was 6.6% as compared to 3.0% precision of entropy. Based on these evaluations, we selected entropy as our preferred measure of cartilage homogeneity.

4.3.2 JSW, Cartilage Volume, Intensity, and Entropy Quantifications

From the automatically segmented cartilage sheets we quantified four different measures: JSW, Cartilage volume, Cartilage intensity, and Cartilage entropy.

As shown in Figure 4.3(a) the JSW quantification was not able to separate the healthy group from KL 1 ($P = 0.9$). However, it was able to separate the healthy group from the rest ($P = 0.005$).

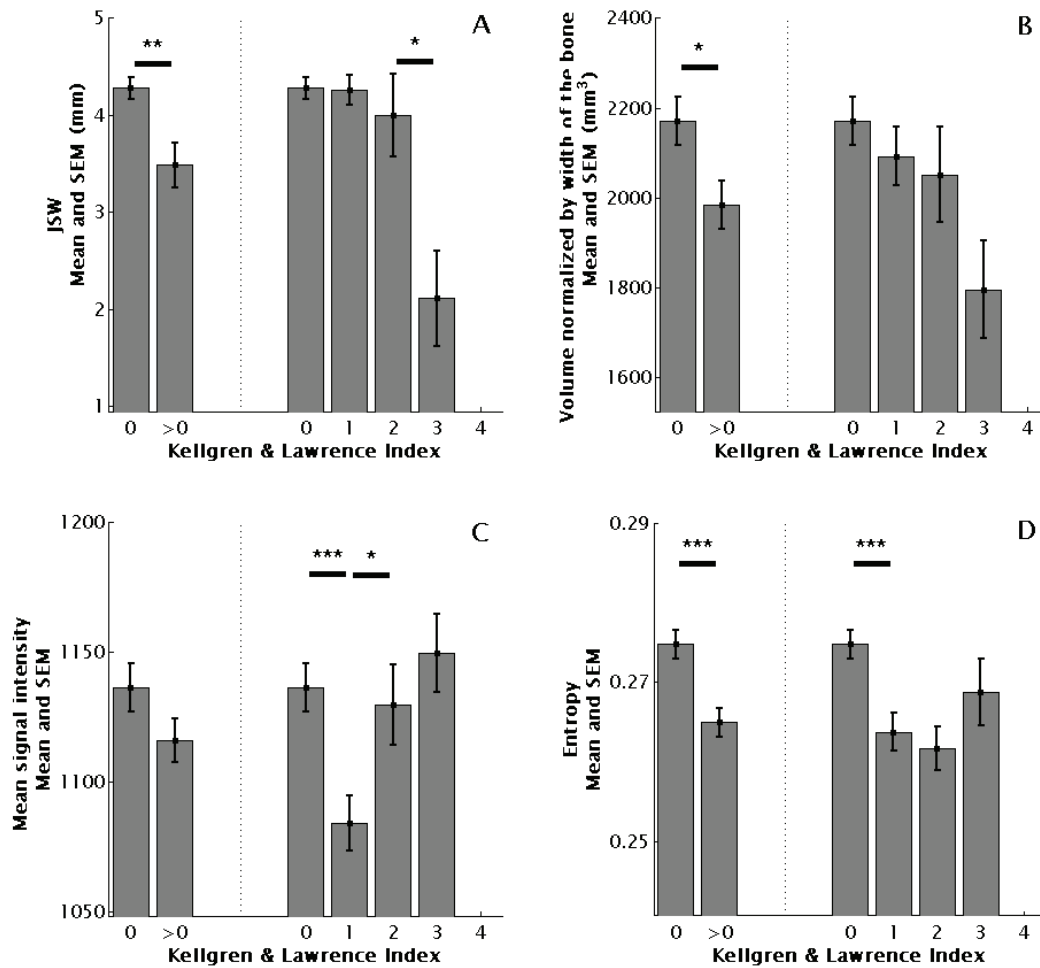


Figure 4.3: Comparison of (a) JSW (b) Volume (c) Mean signal intensity and (d) Entropy as a function of the KL index. Entropy measure can separate healthy group from KL 1 much significantly when compared to the other measures. Moreover it can also separate healthy group from OA.

Traditional cartilage volume was estimated as described in materials and methods. As presented in Figure 4.3(b), volume calculation was not able to significant separate healthy from KL 1. However, a clear and significant trend with a decrease in cartilage volume for increase in KL score was found, as previously published (Folkesson et al., 2005b).

To investigate to which degree the signal intensity level would directly be a suitable biomarker; as a consequence to the strong correlation to water content, we measured the mean cartilage intensity. As shown in Figure 4.3(c), the mean intensity failed to separate the group of healthy from the group of OA subjects. However, mean intensity provided a clear separation of healthy from KL 1 ($P = 0.0009$). The drop from KL 0 to KL 1 followed by a return to KL 0 level for KL 2 and 3 is consistent with the expected increase in water

Table 4.2: P-values for separating KL 0 from KL 1 and KL 0 from KL > 0

Type	JSW	Volume	Mean intensity	Entropy
KL 0 vs KL 1	0.9	0.4	0.0009	0.0004
KL 0 vs KL > 0	0.005	0.02	0.1	0.0003
TEST-RETEST				
Measure	RMS-CV (%)			
JSW	3.9			
Volume	7.4			
Mean signal intensity	3.9			
Entropy	3.0			

concentration in early OA due to swelling and followed by a loss of water content in the later stages.

Lastly, as presented in Figure 4.3(d), the quantification of cartilage homogeneity by entropy was able to clearly separate healthy from OA subjects ($P = 0.0003$). Furthermore, entropy was also able to separate healthy from KL 1 subjects ($P = 0.0004$).

The P -values for the different measures in separating the different groups using the medial tibial compartment are summarized in Table 4.2.

4.3.3 Quantification Precision

The precision of the four quantification methods – JSW, volume, mean intensity, and entropy – are also listed in Table 4.2. The precision of the JSW and automatic cartilage volume quantification were a RMS-CV% of 3.9% and 7.4% respectively. In contrast, the RMS-CV% for mean intensity and entropy were 3.9% and 3.0%, respectively.

4.4 Discussion

We investigated whether intrinsic changes in the articular cartilage related to water distribution visualized by MRI would enable separation of early OA versus healthy individuals. We showed that measurement of cartilage homogeneity quantified by entropy enabled separation of healthy (KL 0) versus OA (KL > 0), and in addition allowed for separation of healthy from early OA (KL 1). Thus, entropy provides sensitivity and information beyond that of JSW, volume, and mean signal intensity quantifications.

Cartilage loss is only secondary to important biochemical changes in the articular cartilage. These changes include protease expression of which Matrix metalloproteinases (MMP) (Sondergaard et al., 2006) and ADAM-TS (Glasson et al., 2005; Stanton H, 2005) are the most well-described pathological activities that result both in proteoglycan depletion and disruption of the collagen network and additionally this process is accompanied by local edema which may be the first signs of decreased cartilage quality.

The population for this study was selected such that it included a large normal population with no or only very minor OA symptoms. This ensured a fairly large number of subjects in the very early stages of OA where only few symptoms are present and thereby allowed an investigation of which cartilage quantifications could detect the early stages with statistical significance.

The present investigation has been based on the use of the KL score as gold standard for estimation of degree of OA. This affects the interpretation of the current MRI data in two important ways. Firstly, it is known that inter-observers differences may result in different annotations of KL score, in particular between KL 0 and KL 1 as well as KL 1 and KL 2 (Altman et al., 1996; Vilalta et al., 2004). Secondly, articular cartilage is not visible by X-ray and only to a smaller extent is accounted for in the annotation of KL scoring compared to that of bone (Hart and Spector, 2003). As a result, the KL annotation may not distribute similar articular cartilage appearances in the same KL group and each KL group may as a consequence include very different presentation of articular cartilage health.

With these difficulties in mind, the present separation of healthy versus KL 1 may seem even more impressive, as undetected articular changes in KL 0 which only are detectable by entropy could have been annotated differently, if the KL scoring system would have been more articular cartilage oriented. Taken together, this impressive identification of early OA changes by automated low field MRI raises the question whether entropy combined with cartilage alterations and subchondral turnover combined with KL 0 and KL 1 would be more appropriate for selecting patients that would benefit from treatment compared to late stages of OA as identified by KL 2 to 4 scoring. Additionally damage to the articular cartilage can be even reversed in early stages before extensive fibrillation, chondrocyte mal-metabolism and differentiation in hypertrophic chondrocytes and apoptosis occurs.

The effectiveness of the homogeneity quantification is related directly to biochemical changes within the cartilage. During the early stages of OA, before cartilage thinning begins, there is a reduction in proteoglycan content and disruption of the collagen framework. Initially this results in increased water content, which leads to swelling of the cartilage (edema). Decreased proteoglycan content is believed to result in increased water content because loss of proteoglycan allows uncoiling of the remaining proteoglycan molecules which increase their negatively charged domain which in turn increases their hydrophilic nature thereby leading to an increase of water (Hollander et al., 1994; Jeffrey and Watt, 2003; McCauley and Disler, 1998). In addition chondrocytes in response to altered material properties tries to compensate by overproduction of proteoglycans (Adams and Brandt, 1991; Braunstein et al., 1990; Myers et al., 1990). On the T_1 MR scans an increase of water leads to a decrease in the average signal intensity as water appears darker than healthy cartilage in T_1 weighted MR (Disler et al., 2000). Thereby

the cartilage at the early stages of OA appears darker, and at the same time more homogeneous.

A few studies have shown that structure of the collagen matrix is responsible for the magnetization transfer effect in articular cartilage (Kim et al., 1993; Seo et al., 1996). Thereby as shown in Figure 4.3(c) in the later stages, as the cartilage begins to disintegrate, the collagen content begins to diminish resulting in a decreased magnetization transfer effect and along with an overall reduction in water content we see a rise in the signal intensity. Therefore the cartilage homogeneity remains high (measured by lower entropy values) for the later stages of OA.

The scanner used in this study is a low-field scanner whereas most recent work is focused on high field MRI. Low field scanners potentially lower costs along with the reduced scan time. Additionally It has been shown that studies done on low-field knee MRI found no clinically significant field-strength-dependent differences in detection of meniscal and ACL tears, and concluded that the performance of low-field MRI equalled that of high-field MRI (Cotten et al., 2000; Ghazinoor et al., 2007). Thereby we expect homogeneity to result in a similar sensitivity on a high-field scanner though a validation study is still needed to reaffirm our claim. Additionally according to (Ghazinoor et al., 2007) the scan resolution used in the this study is sufficient enough for evaluation of the articular cartilage.

Alternative approaches for quantifying cartilage structure from MRI also exist. The T_2 relaxation time is one of the first structural biomarkers based on MRI. It is related to both collagen matrix organization and water content. Lüssea et al showed a correlation between T_2 and water content on 3 subjects in vivo (Lüssea et al., 2000). In recent years, the use of delayed Gadolinium-enhanced MRI of Cartilage (dGEMRIC) to quantify the concentration of glycosaminoglycans (GAGs) in the cartilage has been investigated (Gold et al., 2006). Using 1.5T turbo inversion recovery scans and 15 patients with early OA Tiderius et al showed significant loss of GAG (as measured by the dGEMRIC scores) in the diseased compartment compared to the healthy reference compartment ($P < 0.01$) (Tiderius et al., 2003). However, the dGEMRIC approach is semi-invasive and also more costly than regular MRI for use in clinical trials. Sodium MRI can also be used to measure the reduction of proteoglycan content in the cartilage. Using a test size of 12 patients (9 healthy, 3 early OA) it was shown that Sodium MR imaging has a potential to be used for early detection of OA (Wheaton et al., 2004). A few studies have done early feasibility studies on the use of diffusion tensor MRI for cartilage analysis. In a small in vitro study using a 9.4T scanner, it has been shown that the eigenvector orientations of the local diffusion tensor can separate cartilage layers, which to some degree reflect the alignment of the collagenous fiber network in the articular cartilage (Filidoro et al., 2005). However, due to current limitations in resolution and low signal to noise ratio, only very limited results on in vivo subjects are available.

Compared to these structural quantifications based on sophisticated MR technology, our proposed cartilage homogeneity quantification may allow a detection of early OA with sensitivity and precision comparable to methods based on e.g. dGEMRIC and that too with low-cost equipment (low-field MRI). Furthermore Dray et al argued that T_2 imaging, dGEMRIC and DWI are in their infancy and we still don't have a robust way to monitor collagen changes (Dray et al., 2005).

In conclusion, the use of cartilage homogeneity measured by entropy for detection and quantification of early OA and as an inclusion criterion of patients in longitudinal clinical trials could prove valuable and may thereby eventually aid in improving treatment efficacy.

Separation of Healthy and Early Osteoarthritis by Automatic Quantification of Cartilage Homogeneity from MRI: A Longitudinal Study

Arish A. Qazi, Morten A. Karsdal, Claus Christiansen, and Erik B. Dam. In Proceedings of Imaging Based Measures of OA, 2007.

5.1 Introduction

Future DMOADs should ideally target early OA where it may still be reversible. Therefore, biomarkers that can separate healthy from early OA cross-sectionally as well as longitudinally are required. Cartilage Homogeneity is related to knee cartilage water distribution and altered proteoglycan distribution/integrity and may therefore be a suitable biomarker for very early OA.

5.2 Objective

The aim of this study was to evaluate the performance of the homogeneity measure in detecting the early changes in the knee cartilage over time and comparing it to volume quantification.

5.3 Methods

A population was scanned at baseline and after 21 months using a Turbo 3D T1 sequence (flip angle 40°, TR 50 ms, TE 16 ms, scan time 10 minutes,

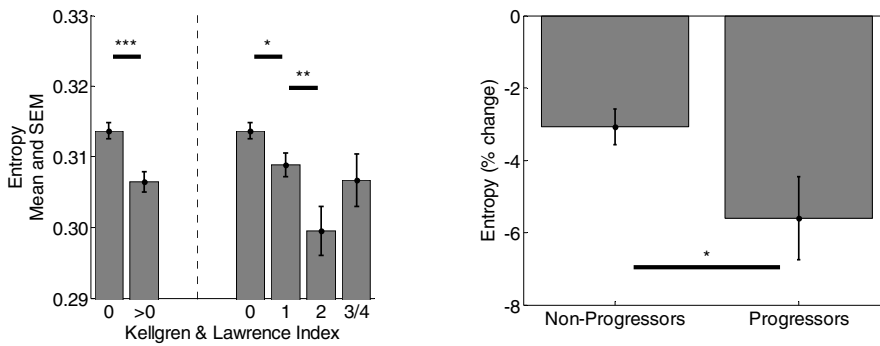


Figure 5.1: (a) Performance of cartilage homogeneity as a function of the KL-index.(b) Performance of cartilage homogeneity in separating the progressors from non-progressors.

resolution 0.7mm*0.7mm*0.8mm) on a 0.18T MRI Esaote scanner. At baseline there were 313 knees of which 25 were used for training of computer-based methods. The validation set had 288 right and left knees at baseline (subject aged: 21-81, females: 44%, BMI: 26.7 ± 4.3) and 243 at follow-up. The knees were examined by radiography and categorized by the Kellgren and Lawrence (KL) Index (with distribution [145, 88, 30, 24, 1] for KL 0 - 4). The medial compartments of the tibial and femoral cartilage sheets were segmented using a fully automatic voxel classification scheme and the cartilage volume and homogeneity were quantified. Homogeneity was quantified by measuring entropy from the MRI signal intensities - this quantifies cartilages with fewer, more dominant intensities as being more homogeneous. For precision evaluation, 31 knees were re-scanned a week after baseline. The healthy subjects at baseline were divided in two groups: 1) 101 subjects that remained healthy at follow-up and 2) 25 subjects that progressed to early OA (KL 1). For each group and both volume and homogeneity changes over the 21 months were computed and the statistical significances based on an un-paired t-test were calculated.

5.4 Results

The scan-rescan precision (mean CV%) of volume and homogeneity were 3.6% and 2.7%. Figure 5.1(a) shows that homogeneity succeeded in separating healthy from early OA ($p < 0.05$). Figure 5.1(b) shows that homogeneity succeeded ($p < 0.05$) in separating the progressors from non-progressors. The decrease in volume (4.0%) for the progressors was less than the decrease in entropy (5.6%) - in particular relative to the respective measurement precision.

Osteoarthritic Cartilage is more Homogeneous than Healthy Cartilage – Identification of a Superior ROI Co-localised with a Major Risk Factor for Osteoarthritis

Arish A. Qazi, Erik B. Dam, Mads Nielsen, Morten A. Karsdal, Paola Pettersen.C, and Claus Christiansen. Academic Radiology, 2007.

Abstract

Cartilage loss as determined either by magnetic resonance imaging (MRI) or joint space narrowing by x-ray is the result of cartilage erosion. However, metabolic processes within the cartilage that later result in cartilage loss, may be a more sensitive assessment method for early changes. Recently it was shown that cartilage homogeneity visualized by MRI representing the biochemical changes undergoing in the cartilage, is a potential marker for early detection of knee osteoarthritis (OA) and is also able to significantly separate groups of healthy subjects from those having OA.

The purpose of this study is twofold. Firstly, we wish to evaluate whether the results on cartilage homogeneity from the previous study can be reproduced using an independent population.

Secondly, based on the homogeneity framework we present an automatic technique that partitions the region of interest in the cartilage that contributes most to discrimination between healthy and OA subjects, and allows for identification of the most implicated areas in early OA. These findings may allow further investigation of whether cartilage homogeneity reveals a predisposition for OA, or whether it evolves as a consequence to disease and thereby can be used as a progression biomarker.

Design: 283 right and left knees from 159 subjects aged 21 to 81 years were scanned using a Turbo 3D T₁ sequence on a 0.18T MRI Esaote scanner. The medial compartment of the tibial cartilage sheet was segmented using a fully automatic voxel classification scheme based on supervised learning. From the segmented cartilage sheet, homogeneity was quantified by measuring entropy from the distribution of signal intensities inside the compartment. Each knee was examined by radiography and the knees were categorized by the Kellgren and Lawrence (KL) Index. Next, based on a gradient descent optimisation technique, the cartilage region that contributed to the maximum statistical significance of homogeneity in separating healthy subjects from the diseased was partitioned. The generalizability of the region was evaluated by testing for over-fitting. Three different regularization techniques were evaluated for reducing over-fitting errors.

Results: The p-values for separating the different groups based on cartilage homogeneity were $P = 2 \times 10^{-5}$ (KL 0 vs KL 1) and $P = 1 \times 10^{-7}$ (KL 0 vs KL > 0). Using the automatic gradient descent technique, the partitioned region was towards the peripheral part of the cartilage sheet. Using this region the p-values for separating the different groups based on homogeneity were $P = 5 \times 10^{-9}$ (KL 0 vs KL 1) and $P = 1 \times 10^{-15}$ (KL 0 vs KL > 0). The precision of homogeneity for the partitioned region assessed as a test-retest root mean square coefficient of variation (RMS-CV %) was 3.3%. Bootstrapping proved to be an effective regularization tool in reducing over-fitting errors.

Conclusion: The validation study supported the use of cartilage homogeneity as a tool for early detection of knee OA and for separating groups of healthy subjects from diseased. Our automatic, unbiased partitioning algorithm based on a general statistical framework outlined the cartilage region of interest that best separated healthy from OA on the basis of homogeneity discrimination. We have shown that OA affects certain areas of the cartilage more distinctly, and these areas are more towards the peripheral region of the cartilage. We propose that this region corresponds anatomically to cartilage covered by the meniscus in healthy subjects. This finding may provide valuable clues in the early detection and monitoring of OA and thus may improve treatment efficacy.

Keywords – Knee Osteoarthritis, Cartilage, Homogeneity, Image Processing, Segmentation, Magnetic Resonance Imaging, Meniscus.

6.1 Introduction

Osteoarthritis (OA) is a degenerative joint disease that is a major cause of disability. Degeneration of the articular cartilage in combination with an altered subchondral compartment are key features of OA (Conaghan et al., 2006; Ding et al., 2005). In terms of quality of life and chronic disability OA is second only to cardiovascular diseases (Haq et al., 2003; McCauley and Disler, 1998). It is estimated that more than one-third of the population above the age of 35 will at some point in their lives experience OA (McCauley and Disler, 1998). At present there are no accepted treatments for OA as no drugs have been consistently shown to modify joint structure or even reverse joint pathology in face of the currently available treatments that are directed towards relief of symptoms (Altman, 2004). Since the prevalence of OA increases with age, the increasing life span will increase the socio-economic impact of OA. Radiographic evidence of knee OA, the most commonly affected weight-bearing joint, can be found in one-third of those above the age of 63 (Felson et al., 1987). Research is on-going to discover disease-modifying anti-osteoarthritis drugs/agents (DMOADs) (Abadie et al., 2004). However to assess the effectiveness of DMOADs, quantification of the structural changes undergoing in the cartilage during the early stages of the disease is needed and for that early diagnosis of OA is essential, both for early identification as well as monitoring of response to treatment in the development of novel interventions for OA.

The currently accepted standard for monitoring progression in knee OA is measurement of the joint space width (JSW) (between the femur and tibia on the knee joint) from radiographs (Altman et al., 1996). This is an indirect evaluation as the cartilage is not visible in X-ray and it is potentially prone to diagnosing the disease relatively late in its course (Calvo et al., 2001). A second technique available for detecting OA is arthroscopy – which even though invasive does not permit visualization of internal cartilage abnormalities. Recently, magnetic resonance imaging (MRI) has received much attention in assessment of the articular cartilage. This can be primarily attributed to the fact that MRI is non-invasive, provides good soft tissue contrast, high spatial resolution and moreover the articular cartilage can be directly visualized and quantified from the MR scans (Eckstein et al., 2006; Pessis et al., 2003). This provides valuable information with regards to morphological and possibly biochemical parameters, that may be associated with the integrity of the articular cartilage. Quantitative measurements from MRI like cartilage volume and thickness measures are now being widely used to monitor progression of OA (Dam, 2006; Folkesson et al., 2007a; Pakin et al., 2002; Solloway et al., 1997).

The pathology of OA involves changes in both the subchondral bone and articular compartments. Presently, late stage detection methods most certainly represent subchondral bone changes as well as cartilage thinning (Buckland-Wright, 2004), as measured by JSW or Kellgren & Lawrence (KL)

score (Kellgren and Lawrence, 1957). Even early OA, as detected by the KL score, may be the result of long ongoing biochemical processes leading to both bone and cartilage alterations. Thus, these techniques may not sufficiently identify patients which may have their cartilage disease attenuated or even reversed.

Cartilage loss is only secondary to important biochemical changes in the articular cartilage. Prior to cartilage fibrillation, local articular edema and swelling are well-recognized features (AlHadlaq et al., 2004; Calvo et al., 2001; Calvo et al., 2004; Maiotti et al., 2000). These intrinsic changes in cartilage water content and spatial distribution are most likely predecessors to the traditional cartilage loss detected by MRI and x-ray based techniques.

Recently, a novel approach for cartilage assessment was introduced, *Cartilage Homogeneity* (Qazi et al., 2006), which was based on the distribution of signal intensities inside the cartilage compartment and was affected by the changes in cartilage water content. Homogeneity performed statistically better than volume quantification in detecting early OA and for separating groups of healthy subjects from those having OA (Qazi et al., 2006). The technique was preliminarily evaluated on 114 knees.

The knee is loaded asymmetrically (Biswal et al., 2002), where the medial compartment of the knee is subject to significantly higher loads relative to the lateral compartment which results in loss of cartilage integrity in the central regions of the medial compartment compared to other regions (Amin et al., 2005; Biswal et al., 2002). Thereby, quantification of cartilage thickness is therefore probably most effectively performed in the central weight-bearing regions in the tibio-femoral joint, as they are subjected to most of the load over the gait cycle (Koo et al., 2005). However, these studies focus on the cartilage loss, which is secondary to the internal biochemical changes.

The purpose of this study is twofold. First, we investigated whether the previous results on cartilage homogeneity were reproducible using an independent, larger population. Secondly, we wished to investigate whether unbiased investigation of internal structure, as measured by cartilage homogeneity, would identify the load-bearing region or alternative regions of interest, as the most dominant, pathological region for discriminating healthy from OA subjects. For this, we present an automatic, un-biased statistical framework for determining the region of interest with the best discriminatory power.

6.2 Materials and Methods

6.2.1 Population

Two populations were used for this study. The first population, consisting of 71 subjects contributing with 114 knees, was used in the previous study where

cartilage homogeneity was introduced (Qazi et al., 2006). The second population, independent of the first population, consisted of 88 subjects with a total of 169 knees.

For the second part of the study – focusing on establishing the most pathological region of interest – the two populations were merged. This gave a total of 159 subjects (82 men and 77 women) from 21 to 81 years old.

Both populations were selected to have a broad range of age as well as OA symptoms from no or only minor OA symptoms to severe symptoms. This composition was designed in order to ensure that a large healthy group could

Table 6.1: Characteristics of the evaluation set (knees = 283, participants = 159)*

Factor	KL 0	KL 1	KL 2	KL 3	KL 4
Number of knees (K)	140	87	31	24	1
Age (yrs)	48.0 ± 16.7	62.7 ± 11.3	65.9 ± 7.4	67.4 ± 4.8	72.7
Females (%)	46%	54%	48%	42%	100%
Weight (kg)	75.3 ± 13.7	76.3 ± 13.6	84.7 ± 14.6	82.5 ± 11.6	84.8
Height (m)	1.73 ± 0.09	1.7 ± 0.09	1.69 ± 0.09	1.7 ± 0.1	1.66
BMI (kg/m ²)	25.1 ± 4.0	26.4 ± 4.1	29.4 ± 3.7	28.8 ± 3.6	30.7

* Values are mean ± SD. KL refers to the Kellgren and Lawrence Index

be compared to both a considerable group with early OA as well as a considerable group with more progressed OA. Additionally, the populations were thereby fairly representative of the general population. Selection criteria ensured that none of the subjects had previous knee joint replacement, inflammatory arthritis in the knee, or any contraindication for performing MRI examination. Subjects underwent both clinical and radiological examinations and were classified into groups of healthy subjects or OA patients according to the ACR (American College of Rheumatology) definition of OA. From the total of 318 knees, 10 were excluded due to poor image quality in either radiographs or MRI. From the remaining 308 knees, 25 were used to train the automatic cartilage segmentation method. So a total of 283 knees (of which 114 were from the previous study, and 169 were from the new population) were used for evaluation of the methods in the following. Details for the 283 knees are given in Table 6.1.

All participants signed approved information consent and the study was carried out in accordance with the principles of the Helsinki Declaration II and European Guidelines for Good Clinical Practice. The study protocol was approved by the local Ethical Committee.

In this study, we focused on the medial compartment of the tibial cartilage since previous studies have shown that OA is most prevalent observed in this compartment (Dunn et al., 2004).

6.2.2 Imaging Protocol

A total of 283 right and left knees were examined and quantified by radiography and MRI. Using radiographs (X-rays) these knees were classified by a radiologist as 0 - 4 on the Kellgren-Lawrence (KL) (Kellgren and Lawrence, 1957) index where KL 0 represents healthy and KL 4 severe OA. The radiologist also marked the most medial and lateral points on the tibial plateau (excluding osteophytes). These points define the width of tibial plateau, which is a measure of the size of the knee.

MRI Image acquisition was done on an Esaote C-Scan low field 0.18T clinical scanner, acquiring a Turbo 3D T_1 sequence (40° flip angle, repetition time 50 ms, echo time 16 ms). The scans were made through the sagittal plane with the resolution of $0.7\text{mm} \times 0.7\text{mm}$ in each slice and a slice thickness between 0.7mm and 0.9mm but typically 0.8mm . The dimensions of the scans were 256×256 pixels with around 110 slices depending on the size of the knee. The approximate scan time was 10 minutes.

For reproducibility evaluation, the protocol was repeated a week later on 31 knees.

6.2.3 Automatic Cartilage Segmentation

The medial compartment of tibial cartilage sheets were automatically segmented using voxel classification based on supervised learning (Folkesson et al., 2007a). For the classification we used an approximate k NN classifier, which was implemented in an ANN (Approximate Nearest Neighbor) framework. To increase the precision of the method, a scheme was incorporated which accounted for variation in the subject's position in the MR scanner (Folkesson, 2006). Furthermore the computational time was significantly reduced by incorporating an efficient classification scheme (Dam et al., 2006a). Figure 4.1(a) and Figure 4.1(b) shows an MR scan slice illustrating the automatic segmentation. From the segmented cartilage sheet, volume for each subject was computed by counting the number of cartilage voxels. The volume measure was normalized by the width of the subject's tibial plateau.

6.2.4 Quantification of Homogeneity

Homogeneity was quantified by measuring entropy from the segmented cartilage sheet (Qazi et al., 2006). Entropy is well-known in information theory (Shannon, 1948) as a measure of information content (or inversely randomness) present in the data. Entropy is measured from the signal intensity histogram, which represents the distribution of intensities present inside the cartilage compartment.

A histogram H can be defined by the following formulation: $H(i) = n_i$; where n_i ($i = 0, 1, \dots, L-1$) represents the number of occurrences for signal intensity i and L represents the number of distinct grey levels in the cartilage. The histogram was normalized by the total number of intensities N such that the histogram represented the probability distribution of the signal intensities – and thereby also became invariant to the cartilage volume.

Furthermore, the intensity range was divided into B equal sized bins. The bin width was chosen to be 100 using the Freedman-Diaconis rule (Freedman and Diaconis, 1981). Using this normalized binned histogram we calculated entropy as:

$$E = - \sum_{i=0}^{N-1} H_i \log(H_i) \quad 6.1$$

Entropy quantifies cartilage with fewer, more dominant intensities as being more homogeneous. In the extreme, the most homogeneous distribution will only have a single intensity present – such a histogram can be described with very little information and therefore has minimal information content and minimal entropy value. In contrast, the higher the entropy value the more heterogeneous the cartilage will be. As an example, Figure 4.2(a) and Figure 4.2(b) shows the histograms of knee cartilage with highest entropy as well as with the lowest entropy in the dataset. When assessing the information content, the specific intensities and their ordering in the histogram are irrelevant. The entropy value depends on whether many in-frequent intensities are present (heterogeneous, high entropy) or whether relatively few frequent intensities are dominant (homogeneous, low entropy). Thereby, unlike a simpler, homogeneity-like measure like intensity standard deviation, entropy is not assuming that the intensity values follow a Gaussian distribution. Due to both the noise properties of MRI (that result in a Rician distribution (Gudbjartsson and Patz, 1995)) and partial volume effects, a Gaussian assumption is wrong.

6.2.5 Statistical Analysis

To evaluate the significance of homogeneity as a biomarker for early detection of OA and for separating the group of healthy subjects from those having OA, the knees were divided in groups based on the KL index; and using the unpaired student t-test, P -values for separating the groups based on entropy were calculated. A 0.05 level was used throughout the paper for statistical significance. The significance levels were computed for entropy measured on medial tibial cartilage sheet and then compared to the significance levels of separating the same groups using the cartilage volume quantification.

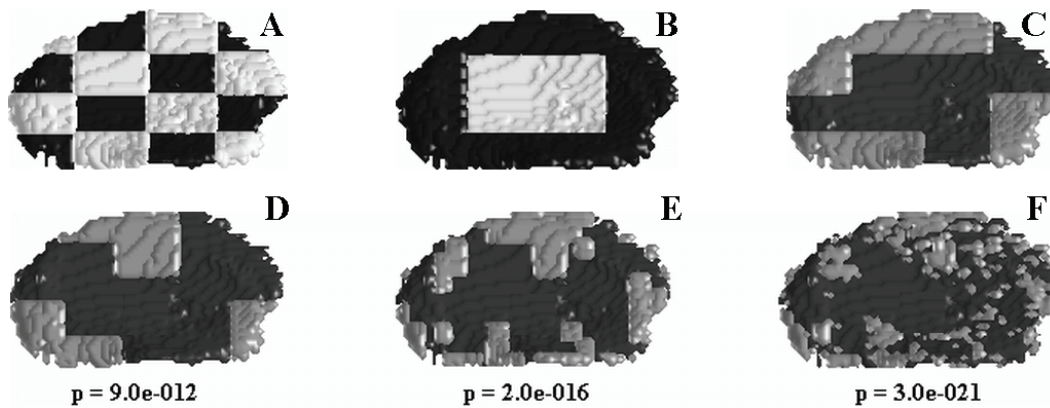


Figure 6.1: Illustration of the partitioning algorithm for resolution 4×4 (a) Cartilage division phase (b) Inserting boundary blocks in the queue (c) Partitioned cartilage. (d-f) Regions partitioned by the algorithm on a sample cartilage under different resolutions. The areas shaded in light gray are the partitioned regions (of high homogeneity discrimination). We can see that these regions are more towards the non-central regions of the cartilage. (d) resolution 5×5 (e) 10×20 (f) 80×140 .

6.2.6 Cartilage Region Partitioning

The quantification of homogeneity and the statistical analysis used for evaluating the significance of discrimination of healthy and OA subjects is described above where the cartilage is considered as one single region. The following presents an algorithm for partitioning the cartilage yielding the region of interest with most statistical, discriminatory power – i.e. the region that best distinguishes healthy from OA subjects. The algorithm is un-biased in the sense that no anatomical knowledge of the surrounding anatomical structures is incorporated.

The partitioning algorithm consists of two phases, a cartilage division phase and a block-shrinking phase. In the division phase each cartilage was sliced in squared blocks (see Figure 6.1(a)) by mapping each cartilage sheet on a rectangular grid where the dimensions of the grid were determined by the height and width of the cartilage sheet. The number of rows and columns determine the size of each block. This grid ensures that a particular block corresponds to approximately the same anatomical area across all the cartilage sheets. As an example, Figure 6.1(a) shows a division of 16 blocks (resolution 4×4) on a sample cartilage.

The block-shrinking phase searches for the region of interest – given by a set of blocks – most indicative of the pathology of the disease. This “importance” of a set of blocks is determined by the significance (p-value) of the separation of healthy from OA subjects based on entropy. In principle, all possible block sets could be tested. However, for small block sizes, the number of possible combinations makes exhaustive evaluation infeasible.

Therefore, the block-shrinking phase starts with the entire cartilage sheet included in the region of interest, and then shrinks the region of interest by excluding those that increase the discriminatory power. The shrinking is done by iteratively evaluating whether the blocks at the periphery of the provisory region of interest should be excluded – this is done by comparing the discriminatory power of the two sets where the block is either included or excluded. Whenever a block is excluded, the periphery of the current region of interest changes, and new blocks are to be inspected. This process is illustrated in Figure 6.1 (a-c) for large blocks. In Figure 6.1(d-f) the results of applying the algorithm with decreasing block size is illustrated.

With small blocks a huge number of possible resulting regions of interest exist. Thereby, there is a risk of over-fitting the solution to the training data – i.e. optimizing the region of interest to fit irrelevant details in the training data that are unlikely to appear in other groups. This is very likely to occur with relatively small training sets. Therefore, the generalizability (the ability to generalize to other populations) ability must be evaluated.

To evaluate the generalization ability of the algorithm, the original set of 283 knees randomly in was divided into two subsets, the training set and the test set, such that the KL distribution of each set resembled the original set. The training set consisted of 141 knees. The partitioning algorithm was evaluated on the training set and then the resulting region of interest was evaluated on the test set. The correspondence between the significance levels of homogeneity (measured by the p-value) for the training and test sets determined the generalizability of the algorithm.

In case of low generalizability, implying over-fitting, this problem can often be solved by regularization of the region of interest. Thereby, the region is enforced to be more regular and simpler – which lowers the ability to model minor (likely irrelevant) details and therefore often increases the generalizability.

Three different regularization approaches were evaluated for their ability to enforce generalizability. The first approach regularized the region using geometric curve evolution by iterating the median filter (Cao, 2003). The second approach utilized binary morphological filtering on the region (Gonzalez and Woods, 2002). Both are standards approaches used in image processing to simplify regions.

The third approach regularized the region of interest in two steps. It utilized a combination of simple binary morphological processing and a statistical methodology known as bootstrapping. Bootstrapping is a simple but powerful Monte Carlo method to assess statistical accuracy or to estimate a distribution from statistics from few samples (Efron and Tibshirani, 1993). For the bootstrap regularization method, the 283 knees were first divided in a test set of 142 knees and a training set of 141 knees. From the training set, a set of

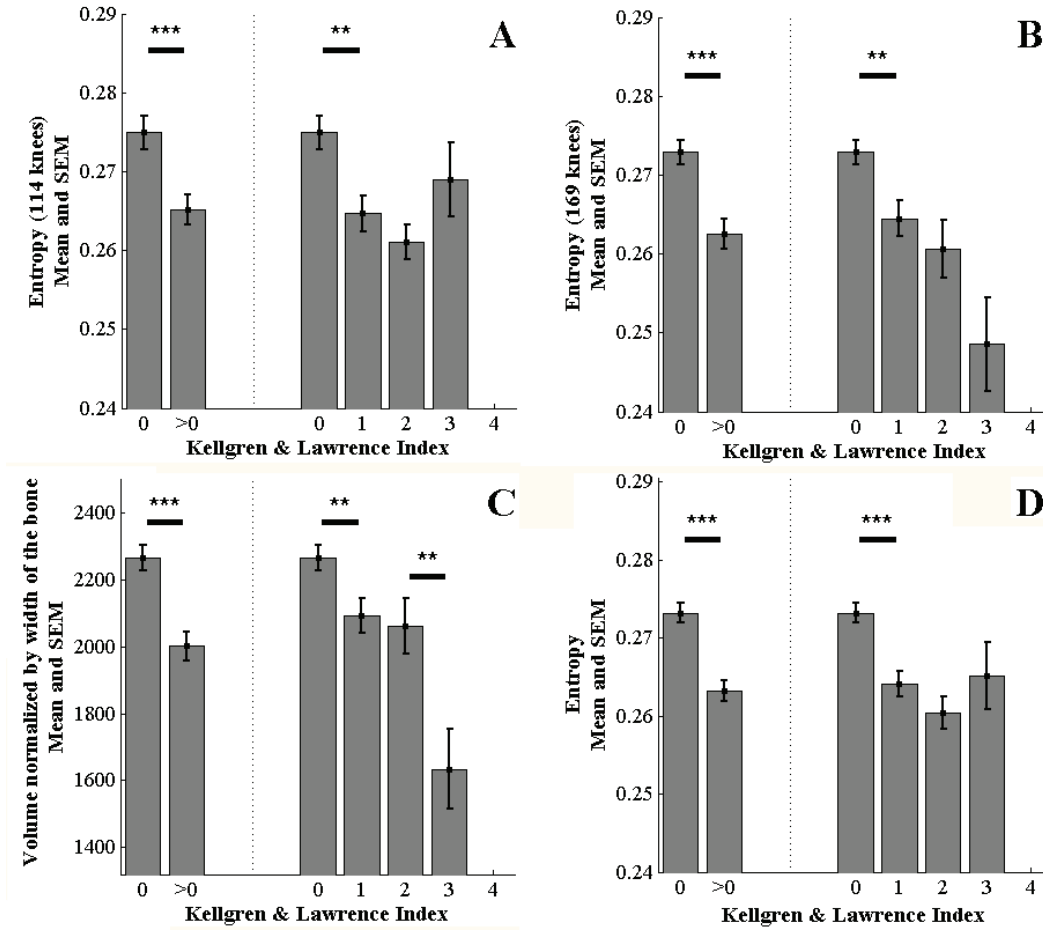


Figure 6.2: Comparison of (a) Entropy: 114 knees and (b) Entropy: 169 knees. Comparison of (c) Volume and (d) Entropy as a function of the KL index for 283 knees. Entropy measure can separate healthy group from KL 1 much significantly then volume quantification. Moreover it can also separate healthy group from OA.

70 knees was sampled while keeping the KL distribution similar to the original set. Then the partitioning algorithm was applied and the resulting region was pre-regularized using morphological filters (closing followed by opening, each with a 3x3 filter) giving a certain region of interest. This step was repeated 1000 times, each time taking a new, random set of 70 knees from the training set. The 1000 resulting regions were added together to form the *Vote Map* – where each block had a number of votes determined by how many of the 1000 trials included that block in the resulting region of interest. This *Vote Map* gives an estimate where the bootstrapping regularization simulates a much larger training set and thereby ensures regularization. The *Vote Map* was then finally thresholded at 500 to give the region of interest included in the majority of the random trials.

The region of interest resulting from the bootstrap regularization was evaluated on the test set of 142 knees. We compared the resulting *p*-value

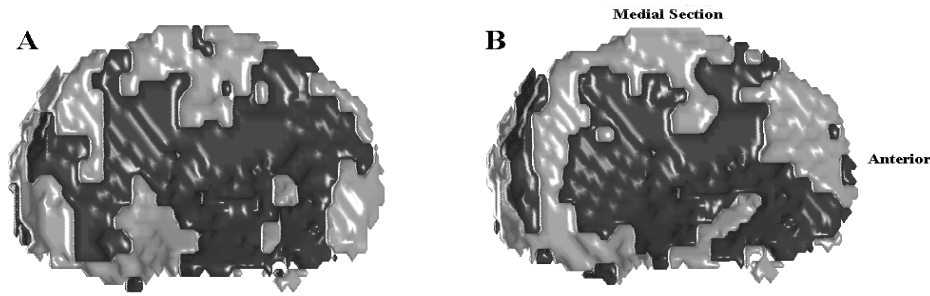


Figure 6.3: (a) and (b) shows the same knee (resolution 80 x 140) but with two different executions of the partitioning algorithm on an independent data set of 141 knees. We can see that after morphological processing the two regions are still different. We need a regularization technique that enables us to have one generalized region that could point with much confidence to the disease modifying regions.

with the significance of not using the partitioning algorithm (using the entire cartilage sheet) on this set.

The bootstrap regularized region of interest was also tested for its ability to discriminate healthy from early OA (KL 0) on the basis of homogeneity. The overall precision of the methods was assessed using the test-retest root mean squared Coefficient of Variation (RMS-CV%) of the entropy values (Gluer et al., 1995). For each subject, the CV is defined as the standard deviation of a series of experiments divided by the mean. The overall CV is expressed as an RMS of the subjects CVs.

6.3 Results

6.3.1 Cartilage Volume and Homogeneity

Figure 6.2(a) shows cartilage homogeneity as a function of the KL index for the old data set of 114 knees and Figure 6.2(b) shows it for the new set of 169 knees. The first two columns of Table 6.2 list the p-values for the two datasets respectively.

From the 283 (combining the above two data sets) automatically segmented un-partitioned cartilage sheets we quantified cartilage volume and cartilage entropy.

Cartilage volume was computed as described in materials and methods. As presented in Figure 6.2(c), the cartilage volume measure was able to separate healthy from KL 1 and healthy from the diseased ($P = 6 \times 10^{-3}$ (KL 0 vs KL 1), $P = 7 \times 10^{-6}$ (KL 0 vs KL > 0)). However, when compared to homogeneity the significance levels were not so impressive.

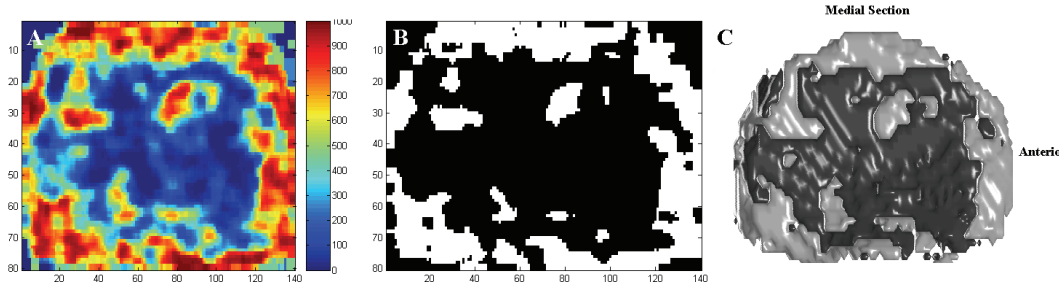


Figure 6.4: (a) Combination of morphological processing and bootstrapping. The *Vote Map* after 1000 randomised trials (b) *Vote Map* thresholded at 500 (c) The same knee from Fig. 3 (resolution 80 x 140) but after regularization.

As presented in Figure 6.2(d), the quantification of cartilage homogeneity by entropy was able to clearly separate healthy from OA subjects ($P = 1 \times 10^{-7}$). Furthermore, entropy was also able to separate healthy from KL 1 subjects ($P = 2 \times 10^{-5}$).

The P -values for the measures in separating the different groups using the medial tibial compartment are summarized in Table 6.2. The precision of automatic cartilage volume quantification was a RMS-CV% of 7.4% respectively. In contrast, the RMS-CV% for entropy was 3.0%.

6.3.2 Cartilage Partitioning

Figure 6.1(d-f) illustrates on a sample tibial medial cartilage the region obtained as a result of executing the partitioning algorithm without regularization on the set of 283 knees using different block sizes. The figure also shows the p-values for each region of interest at the different resolutions. The resolution 80 x 140 was chosen such that the algorithm was executed at sub-voxel level (the mean dimensions of the medial tibial cartilage sheets were 40 x 70 voxels). Table 6.3 summarizes the generalizability results of the algorithm for the test and train sets on different resolutions for three random trails. For large block sizes (5x5 and 10x20) there is no over-fitting. For small blocks (80x140) there is severe over-fitting as the train and test p-values differs by several orders of magnitude.

The results for the first two regularization techniques (morphological

Table 6.2: P-values for separating KL 0 from KL 1 and KL 0 from KL > 0

Type	Entropy (114 knees)	Entropy (169 knees)	Volume (283 knees)	Entropy (283 knees)	Entropy Partitioned (283 knees)
KL 0 vs KL 1	3×10^{-3}	2×10^{-3}	6×10^{-3}	2×10^{-5}	5×10^{-9}
KL 0 vs KL > 0	8×10^{-4}	4×10^{-5}	7×10^{-6}	1×10^{-7}	1×10^{-15}

* Values are mean \pm SD. KL refers to the Kellgren and Lawrence Index

Table 6.3: Generalizability evaluation of the partitioning algorithm
(Table lists the p -values for three different random trials)

Resolution	1		2		3	
	Train	Test	Train	Test	Train	Test
5 x 5	3×10^{-6}	9×10^{-7}	1×10^{-5}	1×10^{-6}	4×10^{-5}	4×10^{-7}
10 x 20	7×10^{-8}	5×10^{-7}	9×10^{-9}	5×10^{-9}	1×10^{-9}	8×10^{-7}
80 x 140	1×10^{-12}	1×10^{-6}	1×10^{-15}	1×10^{-5}	3×10^{-15}	4×10^{-4}
Regularization (80 x 140)	4×10^{-8}	7×10^{-9}	4×10^{-8}	5×10^{-9}	5×10^{-9}	7×10^{-8}

processing and geometric curve evolution) and are not shown as they failed in reducing the train-test error. As an illustration, Fig. Figure 6.3(a-b) shows the partitioned regions obtained from executing the algorithm on two different, non-overlapping, random datasets (141 knees each) and then regularizing them by applying morphological processing. The difference in the resulting regions of interest demonstrates the lack of generalizability.

Figure 6.4 illustrates the results for the third technique of bootstrap regularization. Figure 6.4(a) shows the *Vote Map* for the 80 x 140 resolution and Figure 6.4(b) shows the resulting region of interest (the thresholded map). Figure 6.4(c) also shows the region of interest in a sample knee cartilage sheet. Table 6.3 (last row) demonstrates the generalizability for this regularization method. The resulting p -values for train and test sets are very similar, and demonstrate no over-fitting. Table 6.4 lists the p -values when evaluating the significance of the region on the validation set.

Table 6.4: Bootstrap Regularization using the *Vote Map*
(P -values for the evaluation set of 142 knees)

Type	p -value
Whole cartilage	0.0002
Regularized	0.000009

Finally, Table 6.2 (last column) shows the p -values for separating the different groups when using the regularized region of interest on the full set of 283 knees. The precision of homogeneity for this region assessed, as a test-retest root mean square coefficient of variation (RMS-CV%) was 3.3%.

6.4 Discussion

These data validated the use of the novel parameter “cartilage homogeneity” on an independent population. The results supported that homogeneity, measured by entropy, is superior to other techniques in detecting early OA, as KL1 was significantly separated from KL0 – which is likely not so impressive using traditional techniques such as thickness and volume assessments. We further investigated the potential of the cartilage homogeneity techniques, and identified superior regions of interest that allowed for even better separation of healthy and OA individuals.

The results on the region of interest resulting from the partitioning algorithm demonstrate that – using homogeneity discrimination as the basis – a region of interest with a better, statistical separation of healthy and OA subjects is possible. In order to facilitate a region of sufficient detail to allow any feasible physiological interpretation related to the anatomy, a small block size in the partitioning is needed. However, the generalizability evaluation showed that as the block size went to voxel size or below, over-fitting was a problem. The immense number of possible block set combinations compared to the relatively small population allows adaptation to irrelevant details in the population and thereby ruin the ability to generalize to other populations (as shown in Table 6.3 for the non-regularized 80x140). To solve this problem, a bootstrap regularization scheme was introduced. The results show the regularization succeeds in eliminating over-fitting and ensures generalizability (as shown in Table 6.3, bottom row). The resulting region is illustrated in Figure 6.4(b) and Figure 6.4(c). We can see that the region is relatively simple, smooth and connected in few components.

The results (Table 6.2) show that when applying the region of interest from the bootstrap regularized partitioning (Figure 6.4(c)), the separation of healthy and OA subjects based on homogeneity is stronger than when using the entire sheet ($P = 1 \times 10^{-15}$ compared to $P = 1 \times 10^{-7}$) and far superior to volume quantification ($P = 7 \times 10^{-6}$). Furthermore, even if the region of interest is relatively small, the precision of the quantification is still comparable to the quantification from the full sheet (RMS CV 3.3% compared to 3.0%). For separation of healthy from early OA (KL 1), the bootstrap regularized region of interest is also stronger than when using the full sheet ($P = 5 \times 10^{-9}$ compared to $P = 2 \times 10^{-5}$).

The effectiveness of the homogeneity quantification may be related to biochemical changes within the cartilage. During the early stages of OA, before cartilage destruction in terms on traditional volume assessment, there is altered proteoglycan distribution and disruption of the collagen framework. This is the consequence of chondrocytes which responds to altered material properties by compensation through overproduction of proteoglycans (Adams and Brandt, 1991; Braunstein et al., 1990; Myers et al., 1990), and in part through uncoiling of the remaining proteoglycan molecules which increase their negatively charged domain which in turn increases their hydrophilic nature thereby leading to an increase of water (Hollander et al., 1994; Jeffrey and Watt, 2003; McCauley and Disler, 1998) – and thereby swelling of the cartilage (edema). On the T₁ MR scans an increase of water leads to a decrease in the average signal intensity as water appears darker than healthy cartilage in T₁ weighted MR (Disler et al., 2000). Thereby the cartilage at the early stages of OA (KL 1) appears darker, and at the same time more homogeneous (having lower entropy). Thus cartilage loss is only secondary to important biochemical changes in the articular cartilage.

Previous studies have shown that during OA the cartilage is not affected uniformly. In one study it was shown that cartilage lesions located in the central region of the medial compartment are prone to more rapid progression of cartilage loss than cartilage lesions in the anterior and posterior portions of the medial compartment (Biswal et al., 2002). The region of interest resulting from the bootstrap regularized partitioning based on homogeneity results was far from corresponding to the central, load-bearing part of the cartilage - rather it outlines the peripheral part of the cartilage. We believe that the region that is being partitioned corresponds to the cartilage covered by the meniscus (Figure 6.4(c)). The menisci are wedge shaped cartilages that are interposed between the femoral condyles and the tibial plateau. The meniscus helps in stabilizing and assisting in the distribution of large loads across the joint (Messner and Gao, 1998). Several studies have highlighted the importance of the meniscus in the development of OA. It has been shown that damage and loss to the meniscus affects the articular cartilage as shown by the increased risk of developing OA after meniscectomy (Bennett and Buckland-Wright, 2002). Furthermore, it was recently shown that meniscal tears and extrusions appear to be associated with progression of knee osteoarthritis (Berthiaume et al., 2005). Additionally Hunter et al have shown a strong association between meniscal tears and cartilage loss (Hunter et al., 2006). It has also been shown that there is considerable difference in the mechanical properties, e.g. load bearing, of the cartilage covered by the meniscus and the central region not covered by it (Thambyah 2006).

Our automatic approach partitioned the region on the cartilage based on the only condition that the region was statistically most significant in separating groups of healthy subjects versus OA based on homogeneity. It is imperative to note that the partitioning was done without any prior knowledge of the meniscus. Thereby our results support the fact that the meniscal region might be informative when evaluating the disease. Furthermore this region might also be informative when monitoring early OA, as the region has proved to be quite significant when separating groups of healthy subjects from early OA (Table 6.2). It should be noted here that the region was partitioned while minimizing the p-value for separating groups of healthy subjects from those having OA, therefore the fact that it could also significantly separate early OA further highlights the importance of the region.

Highly interesting, this cross sectional study does not rule out that increased homogeneity is a genetic disposition for early OA. This observation is further nurtured by the fact that entropy seem to be indistinguishable between KL1 through 3, albeit they all were significantly different from the group of healthy. This important question may be related to many assessment techniques by MRI with more focus on cartilage quality than traditional cartilage loss estimated by volume interpretations, and may thereby be related to intrinsic properties of cartilage quality and predisposition for OA. Longitudinal studies are needed to validate this novel homogeneity as a

measure for early OA, in which the possible progression needs to be established.

In conclusion, the use of cartilage homogeneity measured by entropy for detection and quantification of early OA and as an inclusion criterion of patients in longitudinal clinical trials could prove valuable and may thereby eventually aid in improving treatment efficacy. Probing the meniscal region might provide valuable clues in the prevention, early detection and treatment of OA.

A Variational Method for Automatic Localization of the most Pathological ROI in the Knee Cartilage

Arish A. Qazi, Erik B. Dam, Marco Loog, Mads Nielsen, Lauze.F. and Claus Christiansen. In Proceedings of SPIE Medical Imaging, 2008.

Abstract

Osteoarthritis (OA) is a degenerative joint disease characterized by degradation of the articular cartilage, and is a major cause of disability. At present, there is no cure for OA and currently available treatments are directed towards relief of symptoms. Recently it was shown that cartilage homogeneity visualized by MRI and representing the biochemical changes undergoing in the cartilage is a potential marker for early detection of knee OA. In this paper, based on homogeneity we present an automatic technique, embedded in a variational framework, for localization of a region of interest in the knee cartilage that best indicates where the pathology of the disease is dominant. The technique is evaluated on 283 knee MR scans. We show that OA affects certain areas of the cartilage more distinctly, and these are more towards the peripheral region of the cartilage. We propose that this region in the cartilage corresponds anatomically to the area covered by the meniscus in healthy subjects. This finding may provide valuable clues in the pathology and the etiology of OA and thereby may improve treatment efficacy. Moreover our method is generic and may be applied to other organs as well.

Keywords – Image Processing, Segmentation, Level-Sets, Osteoarthritis, Knee Cartilage, MRI.

7.1 Introduction

Osteoarthritis (OA) is a degenerative joint disease that is a major cause of disability. Degeneration of the articular cartilage in combination with an altered subchondral compartment are key features of OA (Conaghan et al., 2006; Ding et al., 2005). At present there is no cure for OA as no drugs have been consistently shown to modify joint structure or even reverse joint pathology in face of the currently available treatments, that are directed towards relief of symptoms (Altman, 2004). Research is on going to discover disease-modifying anti-osteoarthritis drugs/agents (DMOADs) (Abadie et al., 2004). However to assess the effectiveness of DMOADs we need to monitor and quantify the structural changes undergoing in the cartilage.

The current accepted standard for diagnosing knee OA and monitoring progression is measurement of the joint space width (JSW) (between the femur and tibia on the knee joint) from radiographs (Altman et al., 1996). This is an indirect evaluation as the cartilage is not visible in x-rays and is also potentially prone to diagnosing the disease relatively late in its course (Calvo et al., 2001). Recently, magnetic resonance imaging (MRI) has received much attention in assessment of the articular cartilage. This can be primarily attributed to the fact that MRI is non-invasive, provides excellent soft tissue contrast, high spatial resolution and moreover the articular cartilage can be directly visualized and quantified non-invasively from the MR scans (Eckstein et al., 2006; Pessis et al., 2003). Quantitative measurements from MRI like cartilage volume and thickness measures are now being widely used to monitor progression of OA (Dam, 2006; Folkesson et al., 2007a; Pakin et al., 2002; Solloway et al., 1997).

However, cartilage loss is only secondary to important biochemical changes in the articular cartilage (AlHadlaq et al., 2004; Calvo et al., 2001; Calvo et al., 2004; Maiotti et al., 2000). Recently, a novel approach for cartilage assessment from MRI was introduced, *Cartilage Homogeneity* (Qazi et al., 2007c), which was based on the distribution of signal intensities inside the cartilage compartment and was affected by the changes in cartilage water content. Homogeneity performed statistically better than other measures specifically volume quantification in detecting early OA and, for separating groups of healthy subjects from those having OA (Qazi et al., 2006). The technique was preliminarily evaluated on 114 knees.

The knee is loaded asymmetrically (Biswal et al., 2002), where the medial compartment of the knee is subject to significantly higher loads relative to the lateral compartment which results in loss of cartilage integrity in the central regions of the medial compartment compared to other regions (Amin et al., 2005; Biswal et al., 2002). Thereby, quantification of cartilage thickness is therefore probably most effectively performed in the central weight-bearing regions in the tibio-femoral joint, as they are subjected to most of the load

over the gait cycle (Koo et al., 2005). However, these studies focus on the cartilage loss, which is secondary to the internal biochemical changes.

In this paper, we investigate whether internal structure, as measured by cartilage homogeneity, would identify the load-bearing region or alternative regions of interest, as the most dominant, pathological region for discriminating healthy from OA subjects. For this purpose as a first step we present an un-biased statistical framework for localization of a ROI. Next, for the ROI to be simple and anatomically plausible it is then embedded in a variational setting to result in a region that represents the most implicated regions in the cartilage.

To our knowledge there has been no study made to date, which involves a fully automatic system for localization of a ROI in the knee cartilage. However variational techniques are quite popular in image segmentation. Pioneering works in this field include (Chan and Vese, 2001b; Mumford and Shah, 1985; Zhu and Yuille, 1996) notably the work by Mumford and Shah. In recent years variational segmentation techniques have often been based on level sets (Osher and Sethian, 1988) which offer many advantages, among others the implicit representation of regions and their contours. For the same reason the framework presented in this paper will make use of level sets.

7.2 Materials and Methods

This section outlines the data acquisition protocol, followed by quantification of cartilage homogeneity and the method for localization of a ROI.

7.2.1 Imaging and Cartilage Segmentation

For details refer to Chapter 6, Section 6.2.

7.2.2 Localizing the ROI

Localization of the ROI is a multi-step process where each of the individual steps is outlined below:

7.2.2.1 Partitioning the Knee Cartilages

In this step, each cartilage sheet is partitioned in squared blocks along the axial plane. A prerequisite to this step is that all cartilages must be aligned. This is achieved by mapping each cartilage on a rectangular grid where the dimensions of the grid are determined by the height and width of the cartilage sheet. This grid ensures that a particular block corresponds to approximately the same anatomical area across all the cartilages.

This step can be thought as stacking the cartilage sheets on top of each other vertically and then cutting them from top to bottom. In principle the cartilages could also be aligned using non-rigid registration techniques however due to

the small size of the cartilage (few *mm* thick) and the unknown cartilage loss in diseased subjects this step might not be so trivial, therefore to begin with we settle for a simpler approach.

The size of each block is determined by the desired resolution. As an example, Figure 7.1(a) shows a partition of 16 blocks (resolution 4×4) on a sample medial tibial cartilage sheet.

7.2.2.2 Block-Labeling

The block-labeling step searches for the ROI – given by a set of blocks – most indicative of the pathology of the disease. In this paper, we propose the “importance” of a set of blocks to be determined by the statistical significance of cartilage homogeneity in separating healthy subjects ($KL = 0$) from the diseased ($KL > 0$). The phase starts with the entire cartilage sheet included in the ROI, and then shrinks the ROI by excluding blocks such that the significance of homogeneity increases.

In principle, all possible block sets could be tested. However, for small block sizes, the number of possible combinations makes exhaustive evaluation infeasible.

Therefore to be computationally feasible and to have an un-biased ROI the shrinking is done by randomly picking a block and evaluating whether removing it leads to an increase in the significance of homogeneity. The block is excluded from the ROI whenever the significance increases. The algorithm randomly evaluates the blocks until the number of iterations T has been reached. In this way a block can be evaluated a number of times thus having a more robust and un-biased estimate.

7.2.2.3 Block Significance: Statistical Power Analysis

The significance of a block is determined by the statistical power. Statistical power is a measure of the ability (given as a probability) of a statistical test to reject a false null hypothesis H_0 : that two groups (healthy and diseased subjects in our case) have the same distribution, therefore there is no difference between them (Eng, 2003). Typically the purpose of power analysis is to estimate the number of individuals N (or the sample size) that will be required to give adequate power level when conducting a clinical trial. This is very crucial as a larger sample size implies more burden both in terms of cost and time. Therefore a feature that requires fewer individuals to indicate a statistical difference between groups such as healthy and diseased would be preferred.

Thereby we use the sample size estimate to find the ROI in the cartilage that requires the least number of individuals based on homogeneity. The estimation of sample size depends on the following parameters (assuming that both distributions follow a Gaussian; if not then a non-parametric test can

be used): minimum expected difference between the different groups D (also known as the effect size), estimated measurement variability σ^2 , desired statistical power β , significance criterion α , and whether the test is one-or two-tailed. Usually sample size N is computed with a $\beta = 0.8$, $\alpha = 0.05$, and a two-tailed test (Eng, 2003).

The process is illustrated in Figure 7.1(a-c) for large blocks. Figure 7.1(d-f) illustrates the outcome of the algorithm with decreasing block size. In all the experiments $T = 30000$, this ensures that each block is inspected a number of times in order to determine whether it should be included.

The resolution 80×140 is chosen such that the algorithm is executed at sub-voxel level (the mean dimensions of the medial tibial cartilage sheets are 40×70 voxels).

7.2.2.4 Regularization for Generalization and Plausibility of the ROI

The ROI should bear two characteristics. Firstly, it should generalize to unseen data i.e. it should not fit to irrelevant details of a particular dataset that are unlikely to appear in other groups. Secondly, it should be regular enough to be anatomically plausible.

To evaluate whether the region is generalizable the original set of 283 knees is randomly divided into two subsets, the train set and the test set of 141/142 knees respectively. The algorithm (for a particular block size) is evaluated on the train set and then the resulting ROI is evaluated for a sample size estimate on the test set. This step is repeated a number of times, each time taking a different train-test pair. The variation of sample size on the test set determines how stable or generalizable the solution is. The second characteristic is evaluated more by visual inspection, as it is difficult to evaluate the simplicity and the anatomical plausibility of the region otherwise.

In case the ROI does not adhere to the two characteristics, it is then regularized by embedding it in a variational setting. Thereby, the region is enforced to be simpler and more regular which lowers the ability to model minor (likely irrelevant) details and often increases the generalizability.

Let I be the set of images, Ω (having same dimensions as V) be the regularized ROI and V be a probability distribution map stating the probability of a block being inside the ROI. The probability of finding Ω given V, I is by $P(\Omega | V, I)$. Let the relation $I \rightarrow V \rightarrow \Omega$ hold due to the *Markov Assumption* which states that values in any state are influenced only by the values in the state that directly precedes it. Thus state V implies that we cannot extract more information from the images I than what is already present in V . We assume that V is a probability map which tells us how probable a certain block is in tracking the pathology of the disease. The ROI is regularized by segmenting V in a variational setting. Segmenting V can be

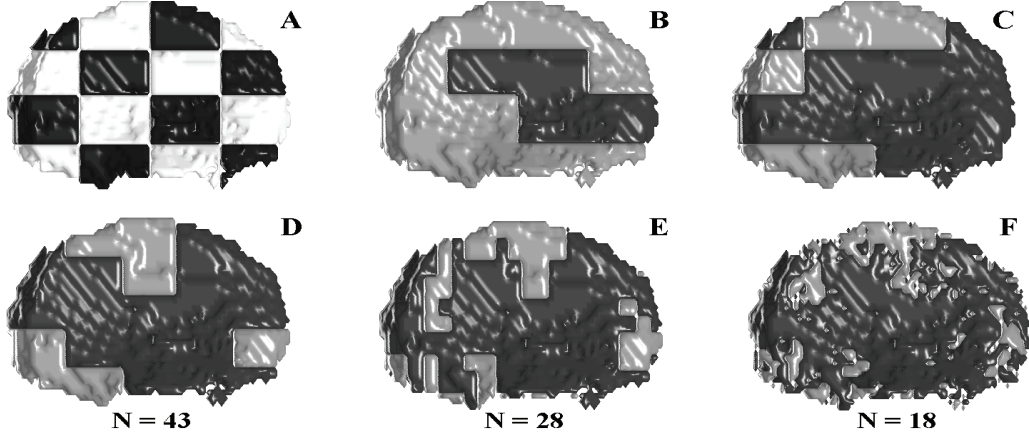


Figure 7.1: Illustration of different steps of the algorithm (4×4). (a) Cartilage partitioning. (b) After 15 iterations of block labeling. (c) After T iterations. (d-f) Un-regularized ROI localized in a sample cartilage under different resolutions from 283 knees. The areas shaded in light gray belong to the ROI. (d) Resolution 5×5 . (e) 10×20 . (f) 80×140 .

expressed in the framework of Bayesian inference by maximizing the conditional probability

$$\arg \max_{\Omega} P(\Omega | V) \propto \frac{P(V | \Omega) P(\Omega)}{P(V)} \quad 7.1$$

The denominator in (7.1) does not depend on the estimated quantities and therefore can be neglected in the maximization. Additionally to be computationally more feasible we assume that the probabilities in V are mutually independent (a commonly used assumption). The first term in the numerator can be written

$$P(V | \Omega) = \prod_{x=1}^D P(V_x | \Omega_x) \quad 7.2$$

where D is the total number of blocks. This conditional probability can be interpreted as the goodness of fit, of Ω to V . Additionally we have

$$P(V_x | \Omega_x) = \begin{cases} V_x & \text{if } \Omega_x = 1 \\ 1 - V_x & \text{if } \Omega_x = 0 \end{cases} \quad 7.3$$

where $\Omega_x = 1$ is a block belonging to the ROI. Let Ω_1 represent these blocks and Ω_2 represent the blocks for which $\Omega_x = 0$. Thus we can write

$$P(V | \Omega) = \prod_{x \in \Omega_1} V_x \prod_{x \in \Omega_2} 1 - V_x \quad 7.4$$

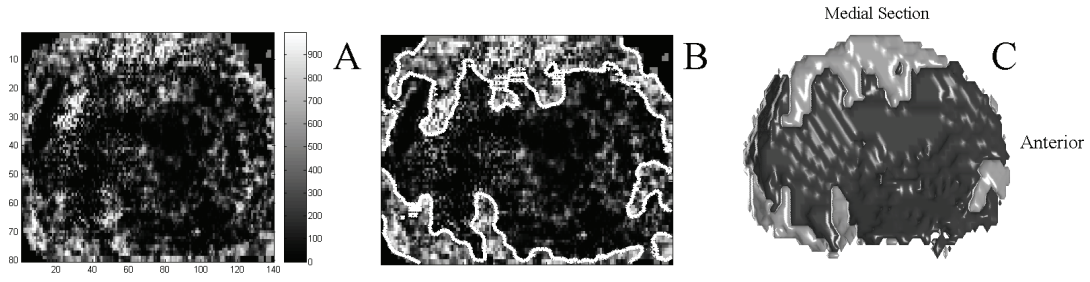


Figure 7.2: (a) Variational regularization: *Vote Map* ($L = 1000$). (b) ROI from the *Vote Map*. (c) Regularized ROI (resolution 80×140) in a sample knee.

We assume the second term in the numerator i.e. the prior probability to only depend on the length of the boundary C separating the two regions (Ω_1, Ω_2)

$$P(\Omega) \propto \exp(-\nu|C|) \quad 7.5$$

Taking the above into account, the *a posteriori* segmentation probability of Ω given the vote map V is determined by

$$\max_{\Omega} P(\Omega | V) = \max_{\Omega, C} \left\{ e^{-\nu|C|} \prod_{x \in \Omega_1} V_x \prod_{x \in \Omega_2} 1 - V_x \right\} \quad 7.6$$

Equivalently, one can minimize the negative logarithm of the above expression which is given by the energy functional

$$E(\Omega, V) = - \int_{\Omega_1} \log(V(x)) dx - \int_{\Omega_2} \log(1 - V(x)) dx + \nu|C| \quad 7.7$$

For minimizing this energy a level set function is introduced. Level sets yield a nice representation for functions and their boundaries and can describe topological changes in the segmentation permitting the splitting and merging of the contour (Osher and Sethian, 1988).

Let boundary C in the functional (9) be represented as the zero level set of a function $\varphi: \Omega \rightarrow \mathbb{R}$ with $\varphi(x) > 0$ for $x \in \Omega_1$ and $\varphi(x) < 0$ for $x \in \Omega_2$. We also introduce the Heaviside function $H(\varphi)$

$$H(\varphi) = \begin{cases} 1, & \varphi \geq 0 \\ 0, & \varphi < 0 \end{cases}$$

The energy can now be written

$$E(\varphi, V) = -\int_{\Omega} H(\varphi) \log(V(x)) dx - \int_{\Omega} (1-H(\varphi)) \log(1-V(x)) dx + \nu \int_{\Omega} |\nabla H(\varphi)|$$

The minimization can now be performed using a gradient descent equation

$$\partial_t \varphi = H'(\varphi) \left(\log \frac{V}{1-V} + \nu \operatorname{div} \left(\frac{\nabla \varphi}{|\nabla \varphi|} \right) \right) \quad 7.8$$

where $H'(\varphi)$ is the derivative of the Heaviside function (also known as the dirac function).

This functional (7.8) is then utilized to localize the regularized ROI from V . The technique above assumes the existence of V . V is computed by employing a statistical methodology known as bootstrap. Bootstrap is a simple but powerful Monte Carlo method to assess statistical accuracy or to estimate the sampling distribution from limited information (Efron and Tibshirani, 1993). For this, the data set is first divided in a train and a test set of 141/142 knees respectively. From the train set, a subset of size K is repeatedly drawn with replacement. Then the algorithm from above is executed on this set and the resulting ROI is stored. This step is repeated L number of times (where L is a large number, typically at least equal to 1000), each time taking a new, random subset from the train set. The L resulting ROI maps are added together to form the *Vote Map* or V – where each location x has a number of votes determined by how many of the L trials included block x .

Figure 7.2(a) shows a sample *Vote Map* ($L = 1000$) which is generated by sampling a subset of $K = 70$ knees from the train set while keeping the KL distribution similar to the original set.

7.3 Results

We begin by evaluating the initial ROI for the two characteristics. Then we show how the variational framework helps in regularizing the ROI.

7.3.1 Generalization and Anatomical Plausibility

Table 7.1 summarizes the generalization results of the algorithm using different block sizes, showing the average and the variation for the randomly sampled train and test sets. The average shows that as we go to finer resolutions we get a better sample size estimate-however the large variation of the test set shows that the ROI is not stable enough to be able to generalize to un-seen data. From visual inspection (Figure 7.1(c-f)) the ROI is not regular enough to be anatomically plausible.

Therefore, to minimize the variation and increase the anatomical correspondence the ROI is regularized.

Table 7.1: Generalizability evaluation of the algorithm without regularization
(The sample size estimates for 100 different random trials.)

Resolution	Train (141 knees) (mean \pm sd)	Test (142 knees) (mean \pm sd)
1 \times 1	93 \pm 49	89 \pm 55
2 \times 1	86 \pm 42	95 \pm 56
3 \times 3	55 \pm 20	75 \pm 38
5 \times 5	43 \pm 15	76 \pm 35
10 \times 20	23 \pm 06	56 \pm 29
20 \times 40	18 \pm 05	56 \pm 28
80 \times 140	14 \pm 04	56 \pm 27

7.3.2 Regularization of the ROI

We chose the 80 \times 140 resolution for regularization as a result of minimum variation on the test set. Figure 7.2 illustrates this regularization technique for a sample train set. Figure 7.2(a) shows the *Vote Map* for the 80 \times 140 resolution and Figure 7.2(b) shows the regularization contour in the *Vote Map*. Fig. 3(c) shows the resulting regularized ROI in a sample knee cartilage sheet. Table 7.2(row 3) demonstrates the generalizability for this regularization method. It shows that with regularization we were able to decrease the sample size estimate of the test set along with the reduced variation. Additionally, from visual inspection the ROI is much more regular than for e.g. from Figure 4.1(f).

Table 7.2: Sample size evaluation of the regularized ROI

Resolution	Train (mean \pm SD)	Test (mean \pm SD)
1 \times 1/Full cartilage	79 \pm 5	76 \pm 5
80 \times 140 (un-regularized)	13 \pm 3	55 \pm 13
80 \times 140 (regularized)	20 \pm 2	46 \pm 2
Load-Bearing	177 \pm 35	171 \pm 36
Meniscus	69 \pm 13	65 \pm 10

*Due to high CPU requirements the results are shown as an average of 3 different random trials. Calculation of one *Vote Map* (L = 1000, T = 30000) using 22 processors takes approximately 72 hours.

7.4 Conclusions

The results on the region of interest resulting from the localization algorithm demonstrate that – using homogeneity discrimination as the basis – a region of interest with a better, statistical separation of healthy and OA subjects is possible. In order to facilitate a region of sufficient detail to allow any feasible physiological interpretation related to the anatomy, a small block size is needed. However, the immense number of possible block set combinations compared to the relatively small population allows adaptation to irrelevant details in the population and thereby ruins the ability to generalize to other populations (as shown in Table 7.1). To solve this problem, a regularization

scheme is introduced. The results show the regularization succeeds in eliminating over-fitting and ensures generalizability (as shown in Table 7.2, row 3). The resulting region is illustrated in Figure 7.2(b) and Figure 7.2(c). We can see that the region is relatively simple, smooth and connected.

Previous studies have shown that during OA the cartilage is not affected uniformly. In one study it was shown that cartilage lesions located in the central region of the medial compartment are prone to more rapid progression of cartilage loss than cartilage lesions in the anterior and posterior portions of the medial compartment (Biswal et al., 2002). The region of interest resulting from the bootstrap regularized partitioning based on homogeneity results was far from corresponding to the central, load-bearing part of the cartilage - rather it outlines the peripheral part of the cartilage. We believe that the region that is being partitioned corresponds to the cartilage covered by the meniscus (Figure 7.2(c)). The menisci are wedge shaped cartilages that are interposed between the femoral condyles and the tibial plateau. The meniscus helps in stabilizing and assisting in the distribution of large loads across the joint (Messner and Gao, 1998). Several studies have highlighted the importance of the meniscus in the development of OA. It has been shown that damage and loss to the meniscus affects the articular cartilage as shown by the increased risk of developing OA after meniscectomy (Bennett and Buckland-Wright, 2002). Furthermore, it was recently shown that meniscal tears and extrusions appear to be associated with progression of knee osteoarthritis (Berthiaume et al., 2005).

Our automatic approach partitioned the region on the cartilage based on the only condition that the region was statistically most significant in separating groups of healthy subjects versus OA based on homogeneity. It is imperative to note that the partitioning was done without any prior knowledge of the meniscus. Thereby our results support the fact that the meniscal region might be informative when evaluating the disease.

Additionally, Table 7.2 (rows 4-5) shows that if we analyze manual delineations of the central weight bearing and meniscal regions of the cartilage then the latter performs better thereby indicating that the meniscal region is informative. Moreover our approach localizes the region automatically and with a better performance level.

Our technique might also be applicable in delineating or regularizing a ROI in other organs. As a regularizer, for example, it can be utilized for spatial regularization of activation maps acquired from functional Magnetic Resonance Imaging (fMRI). Due to a low signal-to-noise ratio (SNR) in fMRI, the activation map consists of noise in the form of a number of scattered small islands. In recent years a few methods have been developed to regularize the noise present in these activation maps (Ou and Golland, 2005).

However as we will see that in contrast to our regularization approach these methods operate on patient specific activation maps and therefore may not be so general. It might also be interesting to probe our algorithm in localization of a ROI in the brain (for e.g. using annotated brain MR scans and suitable feature(s)) and compare that to the activation maps acquired using fMRI.

In conclusion, the use of the method for finding the most implicated region could prove valuable and may thereby eventually aid in improving treatment efficacy. Future work aims at investigating to what extent our current regularization approach can benefit from incorporating learning techniques from data mining and machine learning into the regularization functional. Possibly this would provide improved means to come to an accurate definition of a biologically relevant ROI.

Identifying Regions of Difference for Sample Size Minimization

Arish A. Qazi, Erik B. Dam, Marco Loog, and Mads Nielsen. [in review].

Abstract

This paper attempts to answer, “where?”, in the context of a disease affecting a human organ. For this purpose, we propose a fully automatic statistical framework for identifying biologically interpretable regions of difference between two groups of biological objects. The method is built upon an energy minimization scheme, which finds the regions by minimizing the sample size required to measure the difference between the two groups. The method has been evaluated on synthetic data and clinical data from knee MRI. Based on quantifications from both morphometric and textural based imaging features we have identified the most pathological regions in the articular cartilage. We show that measurements from the resulting regions are reproducible and are able to generalize to unseen data; we achieve a 19 – 51% reduction in sample size. A reduction in sample size is highly desirable for clinical studies, since this reduces the number of participants, which translates to reduced costs and less patient burden. Additionally, identification of such regions can improve the clinical understanding of biological processes involved in the pathogenesis of a disease.

Keywords – Cartilage, Osteoarthritis, Regions of Difference, Sample-Size Minimization, Magnetic Resonance Imaging.

8.1 Introduction

The identification of most affected regions relating to physiological changes in a human organ can help in capturing the dynamics of the underlying pathology. These changes may be attributed to age, genetics, anatomical differences, or changes relating to pathology. For example, it is known that the articular cartilage undergoes morphometric changes during osteoarthritis (Eckstein et al., 2006). Characterization of where the difference between the healthy and the diseased is most pronounced can offer important clues about the pathophysiology of the disease.

Studies involving identification of regions of differences in human organs have been carried out by various researchers. For instance, the distribution of morphometric changes in the brain caused by genetic, environmental factors or various neurodegenerative diseases has been investigated extensively (Andreasen et al., 1994; Dickerson et al., 2001; Pruessner et al., 2000; Raz et al., 1998; Xu et al., 2000). Similar methods have been developed to analyze the articular cartilage. Hohe et al. observed signal intensity changes in the different regions of the patellar cartilage by dividing it into a number of pre-defined sub-regions (Hohe et al., 2002). In a recent study, Wirth and Eckstein (Wirth and Eckstein, 2008) measured regional cartilage thickness in pre-defined anatomically based regions of interest (ROIs). In a longitudinal study, Blumenkrantz et al. quantified differences in structural parameters of bone and cartilage by manually segmenting them in four distinct compartments (Blumenkrantz et al., 2004).

The majority of these studies, however, rely on predefined ROIs. In fact, ROI-based analysis is the current gold standard, specifically for brain atrophy measurements (Good et al., 2002). These methods, however, are labor-intensive and typically can only focus on a limited number of ROIs. Moreover, they do not easily allow comparison of large subject groups. They are observer dependent and thus may result in low inter/intra observer reliability. Another inherent problem is that most often it is not known which region(s) might be most affected by the disease. Thereby, measurements from such methods may lack both sensitivity and specificity.

In order to reliably and accurately detect structural anomalies and other pathological differences in an un-biased way, new techniques have been developed (Freeborough and Fox, 1998; Yushkevich et al., 2003). Voxel-Based Morphometry (VBM) is one such method. Proposed by Ashburner and Friston (Ashburner and Friston, 2000), VBM is increasingly being used to investigate differences in brain morphology between patient and control groups. The output of the method is a probabilistic map which indicates regions of significant gray matter or white matter concentration differences. VBM is widely being used as a tool to examine changes in brain morphometry during healthy aging (Good et al., 2001) or for various neurological conditions

including Alzheimer's disease, and Semantic Dementia (Baron et al., 2001; Mummery et al., 2000).

Similar to VBM analysis, Dam et al. (Dam et al., 2006b) computed 2D thickness maps of articular cartilage loss from knee MRI. The maps show regions of focal thickness loss in the cartilage.

These techniques, however, are based on voxel-by-voxel statistical comparisons where it is assumed that each voxel represents the same anatomical position across all the images, and Bookstein (Bookstein, 2001) pointed out that imperfect registration might lead to interpreting the results as a characteristic of the disease, while in fact this effect might be caused by misalignment of the images. Although smoothing can help alleviate misregistration, the fundamental voxel correspondence problem still remains unsolved. Moreover, the voxel-wise analysis ignores the neighborhood of a voxel which may underscore the anatomical relationship of a voxel.

Previously, we reported preliminary results on identification of regions of most pathological differences in the articular cartilage (Qazi et al., 2008; Qazi et al., 2007a). Both methods employ a computationally intensive bootstrapping technique, based on voxel-wise analysis, to identify a ROI, which is further regularized by curve evolution methods. The ROI problem is combinatorial and therefore, finding an optimal solution requires an exhaustive search, which makes the problem computationally intractable.

The paper explores a new methodology, having statistical underpinnings and incorporating prior knowledge of neighborhood relationships between features, for finding regions of most pathological differences between patient groups. We formulate the problem in a smooth optimization scheme that minimizes the sample size required to measure differences between two groups. A reduction in sample size is highly desirable for clinical studies, since it reduces patient burden and costs. Given spatially normalized feature maps of objects belonging to two groups, the output of the method is a weight map, where each weight reflects the importance of the feature in separation of the two groups. The weight map represents the regions of difference (ROD), where a large weight indicates a region of high difference. The framework is generic and can incorporate any measure for determination of the ROD. Thereby, our framework is able to incorporate both atrophy measures directed towards morphometric differences and textural measures directed towards structural alterations. We evaluate the performance of the framework on both synthetic and clinical data from magnetic resonance (MR) images of the knee.

The objective of the proposed method is to find the optimal feature weights that can discriminate two groups of objects. Linear Discriminant Analysis (LDA) is a standard, well-known machine learning technique used for dimensionality reduction of classification problems by finding a subset of

features that can capture the variability of the two groups (Duda et al., 2001). Therefore, as a good candidate to benchmark against, we compare the performance of our method to LDA.

The paper is organized as follows. In Section 8.2 we present some background information. In Section 8.3 we present the framework for finding the most pathological ROD. In Section 8.4 we analyze the performance of the method on synthetic data. In Section 8.5 we evaluate the method using three different measures for finding the ROD in the cartilage.

8.2 Preliminaries

We start by introducing some necessary concepts concerning the types of errors and sample sizes.

Clinical research is designed to determine if a specific treatment has an effect. Usually this is done by dividing the subjects into two groups, the treatment and placebo and then measuring the effective differences in measurements of a biomarker for the two groups.

When conducting a clinical trial two types of errors must be considered: Type-1 (false positive) and Type-2 (false negative) (Lachin, 1981). A Type-1 error is made when the results of a study indicate a difference between groups when in reality, there is no difference. The probability of a Type-1 error is the p-value and if it is less than a threshold α , typically 0.05, the result is said to be statistically significant, meaning that inferences about a treatment effect based purely on the observed data will be correct.

A Type-2 error occurs when the p-value fails to reach the required level of statistical significance, meaning that there is no observed difference between groups, when in fact there is. The probability of committing a Type-2 error is denoted by β , and its complement $(1 - \beta)$ is known as the statistical power. A common value for power is 0.8 meaning that there is an 80% probability that the difference will be detected.

An important aspect before any clinical study design is the estimation of the required sample size for detecting difference. This is crucial as a larger sample size implies more cost and time along with patient discomfort. The goal of sample size estimation is to reduce the chance of encountering the Type-1 and Type-2 errors. Given desired levels of α , and β . Assuming normally distributed data, sample size N is calculated as (Lachin, 1981)

$$N = \left\lceil \frac{(\sigma_1^2 + \sigma_2^2)(Z_\alpha + Z_\beta)^2}{(\mu_1 - \mu_2)^2} \right\rceil = k \frac{(\sigma_1^2 + \sigma_2^2)}{(\mu_1 - \mu_2)^2} \quad 8.1$$

Where μ_1 and μ_2 are group means, σ_1^2 and σ_2^2 are group variances. Both Z_α and Z_β are probability cut-off points along the x-axis of a standard normal probability distribution. We define $k = (Z_\alpha + Z_\beta)^2$. Equation 8.1 implies that both large differences between groups and smaller variances will reduce the number of participants needed in a trial.

8.3 Methods

This section describes the framework for determination of the ROD. Subsection 8.3.1 formulates finding the optimal ROD problem as an energy minimization scheme. Subsection 8.3.2 and Subsection 8.3.3 present the specifics related to implementation and evaluation of the generalization ability of the method.

8.3.1 Regions of difference Identification by Sample Size Optimization (RISSO)

The method assumes that the biological objects belonging to the two groups have been segmented and are spatially aligned. Additionally, the measure on which the ROD is being computed are known. We note here that our framework: 1) assumes the measure to be Gaussian distributed, and 2) requires that the measure is related to the pathology of the disease in question.

The input to the RISSO framework is a collection of anatomically aligned objects that fall in one of the two groups: G_1, G_2 of size n_1 and n_2 respectively. Each object is represented by an n-dimensional feature map, for instance a 2-D thickness map T . Let the significance of feature i , measured at a given anatomical location, be represented by weight W_i , where $W_i = 0$ means that the corresponding feature is not relevant in separating the two groups. The objective of the framework is to find the optimal n-dimensional weight map W that minimizes the sample size, as given by equation 8.1. The weight map W represents the ROD and can possibly be defined as:

$$W(i) = \begin{cases} 1 & \text{if } i \in \text{ROD}, \\ 0 & \text{otherwise.} \end{cases}$$

For many problems, in order to find the optimal solution, an exhaustive combinatorial search from the 2^q possible subsets of W is required, where q is the number of entries in W . Instead, we propose to define the solution as a smooth optimization problem by relaxing the domain of W to $[0, \infty)$. In this continuous setting we can use variational techniques that will typically allow tractable optimization.

After determining the specific characteristics of the solution we formulate an objective function. From equation (8.1) we can see that devising a scheme that minimizes sample size will equivalently maximize the difference between the groups. The measurements from the two groups are denoted, X_1, X_2 , given as

$$X_1 = [x_1^1, x_1^2, x_1^3 \dots x_1^{n_1}], X_2 = [x_2^1, x_2^2, x_2^3 \dots x_2^{n_2}]$$

Thus, x_j^k , where $j \in \{1, 2\}$, represents a measurement for the k^{th} subject, belonging to one of the two groups. The measurements are assumed to be normally distributed

$$X_1 \sim N(\mu_1, \sigma_1) \text{ and } X_2 \sim N(\mu_2, \sigma_2)$$

We wish to minimize the sample size necessary to distinguish X_1 from X_2 . Hence, given that they can at all be distinguished, we may perform a transformation

$$\widetilde{X}_1 = a_1 X_1 + a_0 \text{ and } \widetilde{X}_2 = a_1 X_2 + a_0$$

$$\text{where } a_1 = \frac{1}{\mu_2 - \mu_1} \text{ and } a_0 = -\mu_1$$

so that $\widetilde{\mu}_1 = 1$ and $\widetilde{\mu}_2 = 0$. Equation (8.1) can be written as

$$N \propto \widetilde{\sigma}_1^2 + \widetilde{\sigma}_2^2, \quad \text{8.2}$$

where $\widetilde{\sigma}_1 = a_1 \sigma_1 + a_0$ and $\widetilde{\sigma}_2 = a_1 \sigma_2 + a_0$

Let $F[W, x_j^k]$ be the value of the measurement of a specific subject, after been weighted by W , and in short denoted as: $F_j^k[W]$. For example, for the 2-D thickness map T , $F = W_i \cdot T_i$ (weighted mean thickness).

Subsequently we have $\widetilde{F} = a_1 F + a_0$ and

$$N \propto \sum_{j \in \{1, 2\}} \frac{\sum_{k \in G_j} (\widetilde{F}_j^k[W] - \widetilde{\mu}_j)^2}{n_j} \quad \text{8.3}$$

where j indicates the two groups. We can formulate the problem as a minimizer for (8.3)

$$\arg \min_{w, a_0, a_1} \sum_{j \in \{1, 2\}} \frac{\sum_{k \in G_j} (\tilde{F}_j^k[W] - \tilde{\mu}_j)^2}{n_j} \quad 8.4$$

where $\tilde{F}_j^k[W]$ is the outcome measure for subject k .

This formulation has an inherent drawback: The functional leads to good separation between groups but does not incorporate prior knowledge about the ROD. In biological settings, the anatomical position of a feature plays an important role and it is likely that neighboring locations are highly correlated. We also prefer a more regularized ROD, which will be less prone to overfitting and can be biologically interpreted as being anatomically plausible.

To exploit the spatial nature of the features and to regularize we add a penalty term of the form $\|\nabla W\|_p$. By ∇ , we denote the gradient operator (or in the discrete setting a local difference operator), such that $\|\nabla W\|_p$ is a measure of the variation or smoothness of W . We select $p = 1$, resulting in $\|\nabla W\|_1$, the L1-norm of the gradient of the weight map. This term is known as zero-order variable fusion and was proposed by Land and Friedman (Land and Friedman, 1997). It has also been adapted for image segmentation by (Chan and Vese, 2001a). The effect of this term is to shrink the solution towards being piece-wise constant. Land and Friedman showed that when compared to other smoothing functionals, such as spline regression, variable fusion produces simpler interpretable solutions and is effective in case of sharp features (Land and Friedman, 1997). An alternate term could be $\|\nabla W\|_2$, however, the term does not produce sparsity in the differences of the features.

Adding the regularization term $\|\nabla W\|_1$ to Equation (8.4)

$$\arg \min_{w, a_0, a_1} \sum_{j \in \{1, 2\}} \frac{\sum_{k \in G_j} (\tilde{F}_j^k[W] - \tilde{\mu}_j)^2}{n_j} + \lambda \|\nabla W\|_1 \quad 8.5$$

The parameter λ influences the extent of spatial regularization. Increasing values of λ will in turn increase the smoothness of the ROD. Assuming that the feature maps relate meaningful anatomical neighborhood relationships, smoothness reflects the local correlation.

In the end, our goal is to choose the optimal features. We prefer a more compact model. A simple model will be easier to interpret and will likely generalize better. Thus, we need a mechanism for feature selection and filtering out the redundant features.

Another motivation for feature selection is that the feature space might typically be much larger than the number of output variables. This is also known as the “large p, small n” paradigm (West, 2003). Such problems may not yield a unique solution but can be solved by eliminating redundant features.

The smoothness term in (8.5) alone is not sufficient since it encourages sparsity in the differences of the features and not on the features. There have been methods proposed for regularization of the solution space, such as ridge regression (Hoerl and Kennard, 1970) and partial least squares (Wold, 1975). A disadvantage of such methods is that the resulting solution space is not very sparse. Proposed by Tibshirani (Tibshirani, 1996), the least absolute shrinkage and selection operator (LASSO) is similar to ridge regression except that it selects the important features while discarding the rest. Therefore, it produces coefficients that are exactly 0, yielding a sparse solution that may be more easily interpretable.

Adding the L_1 -regularization term to functional (8.5)

$$\arg \min_{w, a_0, a_1} \sum_{j \in \{1,2\}} \frac{\sum_{k \in G_j} (\tilde{F}_j^k[W] - \tilde{\mu}_j)^2}{n_j} + \lambda \|\nabla W\|_1 + \eta \|W\|_1 \quad 8.6$$

The parameter $\eta > 0$ controls the sparsity of the solution map W . The regularization terms in our functional are similar to the fused LASSO (Tibshirani et al., 2005), which was applied to 1-dimensional gene expression data.

Furthermore, we impose two conditions on the map W . Firstly, we restrict the weights to being positive, since a negative weight for a feature is not directly interpretable. Secondly, we fix their sum to a constant number; the weights are relative which makes them invariant to scaling. So, to avoid a drift in the optimization scheme and to attain numerical stability we scale the sum of weights to a constant number. The functional changes to a constrained optimization of the form

$$\begin{aligned} \arg \min_{w, a_0, a_1} \sum_{j \in \{1,2\}} \frac{\sum_{k \in G_j} (\tilde{F}_j^k[W] - \tilde{\mu}_j)^2}{n_j} + \lambda \|\nabla W\|_1 + \eta \|W\|_1 \\ \text{subject to} \\ W \geq 0, \sum W = 1 \end{aligned} \quad 8.7$$

Optimizing functional (8.7) will yield a weight map that will identify regions of most discriminate changes between two groups of biological objects.

8.3.2 Optimization of the Functional

The regularization term, $\eta \|W\|_1$, in (8.7) is non-differentiable at $W = 0$, however, since the weights are positive we can utilize smooth gradient based optimization techniques. A number of methods exist, such as the steepest descent, Newton method and quasi-Newton, for a review see (Nocedal and Wright, 1999). In our implementation, because of its efficiency in storage requirements (does not require computation of the Hessian matrix), and convergence rate, non-linear conjugate gradient (CG) descent is utilized to optimize functional (7).

The method is iterative of the form $W_{i+1} = W_i + \alpha_i d_i$, where d_i is the search direction and is computed using the Polak-Ribere (Polak and Ribiere, 1969) update rule, $\alpha_i > 0$ is the step size determined using line search. We implement the line search using Brent's method (Press et al., 2007). We have implemented the method in Matlab.

The functional (8.7) is constrained by the non-negativity constraint on the weights. In order to optimize it as a bound constrained problem we utilize the gradient projection method (Kelley, 1999). Given the current iterate, the weights are projected and scaled to the desired range in order to form the new iterate. The weights are initialized by a constant value such that their sum is one. We experimented with random initializations of the weights and observed that the functional converged to the same minimum but the convergence rate was much slower.

The optimization of functional (8.7) is a two-step process: First, we determine the parameters a_0, a_1 using the current estimate of the weight map. For this purpose we use linear regression. Next, we use these parameters to estimate \tilde{F} followed by computing the regularization terms which finally leads to estimation of the functional.

8.3.3 Finding the Optimal Regularization Parameters

The parameters λ and η have an impact on the characteristics of the weight map and therefore their choice is crucial to the generalization ability of the ROD. In this paper, their values are selected using cross-validation. In cross-validation the data set s is divided into equal parts of $|s|/K$ observations, where K is the number of subsets. The parameters are then trained using $K-1$ subsets and the accuracy of the resultant features is then tested on the K^{th} set.

We choose $K = 3$. The parameters are selected by an optimization framework. Given an initial guess of λ and η , they are optimized by minimizing the sample size on set 2. During optimization, for each specific value of λ and η , the sample size on set 2 is computed by evaluating the weight map estimated

Algorithm 1: Finding the Optimal ROD

```
split the data in 3 sets: S1, S2, and S3
initialize guess vectors for  $\lambda$  and  $\eta$ 
for each element in the  $\lambda$  guess vector
  for each element in the  $\eta$  guess vector
    optimize using CG for N on S2, by varying  $\lambda$  and  $\eta$  on S1
    store the returned optimized parameters and function value
  end
end
choose  $\lambda$  and  $\eta$  corresponding to the minimum function value
use these parameters to optimize the weight map on S1
evaluate the resultant weight map on S3
```

by optimizing functional (8.7) on set 1. Therefore, there are two different optimizations taking place; optimization of λ and η and optimization of the weight map. Both optimizations are carried out using the non-linear optimization framework discussed previously. The optimization is complex and may get stuck in local minima. In order to have stable estimates we use multiple guess vectors for λ and η . The parameters that correspond to the minimum sample size on set 2 are chosen (see Algorithm 1). The weight map corresponding to the optimal parameters, λ and η , is then evaluated on set 3. The reduction of sample size on set 3, when compared to the sample size computed from a weight map with equal weights, or a “uniform weight map”, indicates whether the method is successful.

8.3.4 LDA

The LDA is a well-known scheme for feature extraction and dimensionality reduction (Duda et al., 2001). LDA projects the data onto a lower-dimensional vector space such that the ratio of the between-group distance to the within-group distance is maximized, leading to maximum discrimination between the two groups. We compare the results of our method to those obtained from the LDA. In order to have a fair comparison to our method, we also compare the results to regularized LDA (Friedman, 1989), by regularizing the covariance matrix. The regularization parameter is optimized using cross-validation, similar to the method listed in Algorithm 1, used for optimizing the parameters of our method.

8.4 Results on Synthetic Data

8.4.1 Construction

In order to investigate whether the RISSO framework is able to detect localized regions of differences and is able to generalize, we start with an investigation of synthetic example. The synthetic data consists of 2-dimensional (10×10 pixels) feature maps, belonging to one of two different groups (denoted by G_1 and G_2), with 200 features maps in each group. The G_1 maps are constructed by randomly sampling features from a Gaussian

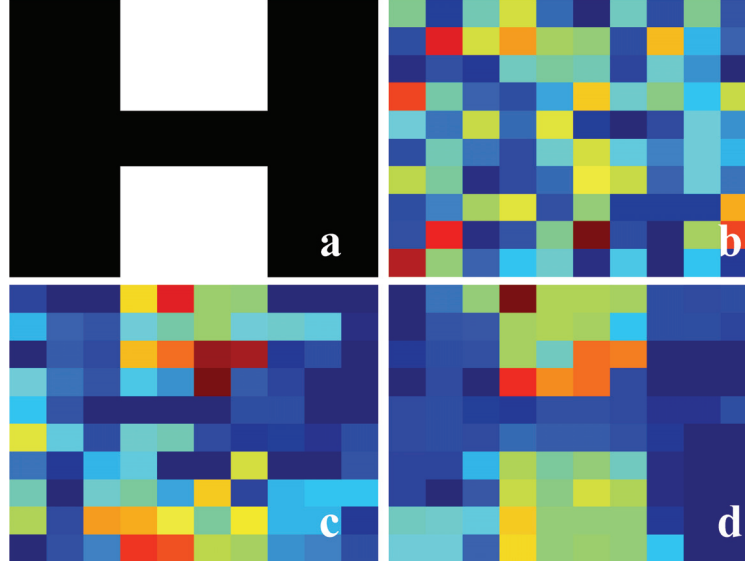


Figure 8.1: Simulated data. (a) Map, without noise, depicting the ground truth ROD (white). (b) G2 map at an SNR of 0.4. (c) ROD, at an SNR of 0.4, determined by Algorithm 1. (d) ROD, at an SNR of 0.4, after increasing the number of objects.

distribution with mean 1. The G_2 maps are similar in construction except that for predefined regions the features are sampled from a Gaussian distribution with mean 0. These regions, illustrated in Figure 8.1a, represent the “ground truth”, ROD.

The feature maps are subjected to additive Gaussian noise. The intensity of the noise is varied to simulate different levels of signal-to-noise ratio (SNR), which is a measure of the quality of image acquisition and feature extraction. We define SNR as the ratio of the mean difference between groups divided by the standard deviation σ of the noise ($1/\sigma$ in our case). As an example, Figure 8.1b illustrates a G_2 feature map at an SNR of 0.4. The feature maps are subjected to Algorithm 1 to validate if the method is able to reconstruct the ground truth ROD, as in Figure 8.1a, and is able to generalize.

8.4.2 Evaluation

Table 8.1 lists the results obtained at different SNR levels. For each SNR the corresponding sample sizes for the ground truth ROD and the optimal ROD, for all the 3 sets are listed. For set 1, when compared to the ground truth sample size, the sample size estimates for the optimal ROD are lower and differ considerably. This implies over-fitting.

For set 2 and set 3 the sample sizes from the optimal ROD are slightly higher than the ground truth sample sizes, however, more importantly the behaviour of sample size estimates for both sets is similar. This implies that apart from a reasonable reconstruction of ROD the method is able to generalize.

Table 8.1: SNR vs. Sample Size on Simulated Data*

SNR	Set 1 (GT)	Set 1 (ROD)	Set 1 (LDA)	Set 1 (R-LDA)	Set 2 (GT)	Set 2 (ROD)	Set 2 (LDA)	Set 2 (R-LDA)	Set 3 (UWM)	Set 3 (GT)	Set 3 (ROD)	Set 3 (LDA)	Set 3 (R-LDA)
2.0	0.42	0.38	0.3	0.44	0.43	0.47	0.65	0.44	1.66	0.38	0.40	0.53	0.41
1.0	3.5	2.8	2.3	2.9	3.8	4.4	6.1	4.5	12	3.1	3.7	4.8	3.7
0.6	14	10	7	7	15	18	26	28	41	12	16	23	25
0.4	88	33	18	18	84	102	201	201	221	52	106	255	255
0.4	103	77	58	60	79	99	159	148	221	74	92	131	127

*GT is ground truth sample size. These have been calculated with the weights set to the binary, ground truth ROD, as shown in Figure 8.1a. UWM is sample size computed from a uniform weight map. R-LDA is regularized LDA.

At a low SNR of 0.4 (Table 8.1, Figure 8.1b), however, the sample sizes from the optimal ROD (Figure 8.1c) are quite different when compared to those of the ground truth ROD. At a very low SNR the data available is not enough to discriminate the difference between the groups. To test this, we experimented by increasing the number of objects from 200 to 500 in each group. We were able to reduce the difference in sample sizes (last row of Table 8.1) between the ground truth and optimal ROD (Figure 8.1d).

To appreciate the method, for set 3, we also list the sample sizes computed from a uniform weight map (UWM). These sample sizes are computed based on a weight map with equal weights for all the features. We can see that when compared to the sample sizes for the UWM, the sample sizes for the optimal ROD are much lower. This implies that the method is able to find a ROD that leads to a better separation of the groups, and is also able to generalize.

The table also lists the sample sizes, computed using the LDA method. At a high SNR, the performance of LDA is similar to our method; however, as the SNR decreases (0.6 and lower) the performance of LDA goes down rapidly. At an SNR of 0.4, for example, the difference between the ground truth sample size and the sample size from the optimal ROD, for both, the LDA and its regularized version (R-LDA), is much larger when compared to our method. Moreover, the sample sizes for the optimal ROD, computed using the LDA, are higher than those obtained for the UWM.

8.5 Results on Clinical Data

The second evaluation of the proposed method is on clinical MR data of the tibial cartilage, which has been used in the study of Osteoarthritis (OA). OA is a degenerative joint disease that is characterized by degeneration of the articular cartilage (Ding, 2005) for which, at present, there is no effective cure and treatment is directed towards symptom relief.

Subsection 8.5.2 outlines the method for alignment of the cartilages. Subsection 8.5.3 outlines how the features are quantified from the aligned cartilages. These are the pre-requisites to the RISSO framework. Subsection 8.5.5 presents the results from our method. Subsection 8.5.6 presents a comparison of the results from our method to those obtained from LDA.

8.5.1 Data Characteristics

A total of 286 knees (both left and right) from 159 subjects (82 men and 77 women, age from 21 to 81 years old) were available (after exclusion of knees used for training the cartilage segmentation method). The knees were examined and quantified by radiography and MRI. Using radiographs (X-rays) these knees were classified by a radiologist as 0 - 4 on the Kellgren-Lawrence (KL) (Kellgren and Lawrence, 1957) index where KL 0 represents healthy and KL 4 severe OA. The number of subjects within each group was: KL0 (144), KL1 (88), KL2 (29), KL3 (24), and KL4 (1). MRI Image acquisition was done on an Esaote C-Scan low field 0.18T scanner, acquiring a Turbo 3D T_1 sequence (40° flip angle, $T_R = 50\text{ms}$, $T_E = 16\text{ms}$).

The scan resolution was $0.7\text{mm} \times 0.7\text{mm}$ with a slice thickness between $0.7 - 0.9\text{mm}$. The dimensions of the scans were 256×256 pixels with around 110 slices. For more details refer to (Dam et al., 2007b). For reproducibility evaluation the MR protocol was repeated a week later on 31 knees.

8.5.2 Cartilage Segmentation and Alignment

The cartilages were segmented based on a supervised voxel classification scheme using a kNN classifier (Folkesson et al., 2007b). The sheets were augmented by an anatomical coordinate system by fitting a deformable m-rep shape model to the segmented articular cartilage (Dam et al., 2007a). The m-rep represents an object by a set of medial atoms, each associated with a position, radius, and directions to the boundary (Pizer et al., 2003).

8.5.3 Feature Measurements

In our experiments, we evaluated three measures: The first two relate to structural changes during the disease. For the articular cartilage these are the *Cartilage Volume* (Folkesson et al., 2007b), which determines cartilage loss, and *Cartilage Thickness* (Dam et al., 2006b), which may be particularly advantageous for the analysis of conditions related to focal thinning of the cartilage. To be invariant to the size of a subjects' joint, both measures were normalized by width of the tibial plateau.

The third measure relates to structure of biochemical changes undergoing in the cartilage. This is *Cartilage Homogeneity* (Qazi et al., 2007b) which relates to intrinsic changes in the articular cartilage water distribution, as visualized by MRI. Homogeneity was quantified by measuring entropy. Entropy is defined

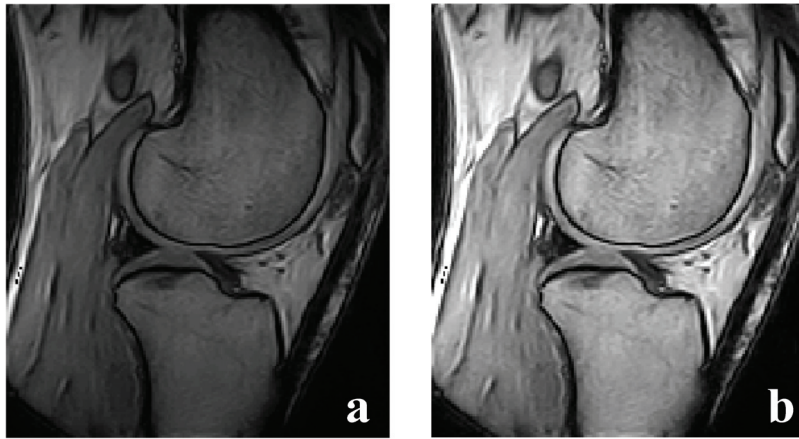


Figure 8.2: Slice from a knee MR illustrating inhomogeneity correction by HUM. (a) before inhomogeneity correction. (b) corrected image.

to be a measure of randomness present in the data. The computation of homogeneity involves signal intensities therefore the scans were corrected for any intensity bias by a method proposed by Guillemaud (Guillemaud, 1998), which filters the image using homomorphic unsharp masking (Brinkmann et al., 1998a).

To quantify feature maps each cartilage was divided in sub-regions, defined by the shape model, which extended from the surface of the cartilage to the cartilage-bone interface. The measurements were then computed for each sub-region and projected onto a 2-dimensional representation of the cartilage. This resulted in a 2D feature map of measurements, where measurements from each sub-region are a feature.

8.5.4 Intensity Non-uniformity Correction

(Note that due to space constraints this specific section is omitted from the submitted version of the paper).

A common MR image model assumes that the intensity values of the voxels of any given tissue class should be constant throughout. MR image data, however, is often corrupted by Radio Frequency (RF) field inhomogeneities, also known as the “bias field”. The presence of such an artifact can significantly reduce the accuracy of segmentation algorithms that rely on absolute pixel intensities. Thus, the intensity nonuniformity correction (NUC) is imperative for quantitative analysis of MR data. Various retrospective techniques have been proposed, for a review see (Hou, 2006).

A common assumption is that the inhomogeneity field is multiplicative in nature (Wells et al., 1996) and consists of low-frequency components (Axel et al., 1987; Johnston et al., 1996). Mathematically this can be represented as follows

$$I_{measured}(X) = I_{corrected}(X) \times B(X) + N(X) \quad 8.8$$

Where $I_{measured}(X)$ is the given image at voxel X , $B(X)$ is the true slowly varying bias field, and $N(X)$ is the noise. The goal of any NUC method is to estimate $I_{corrected}(X)$.

The most intuitive approach from (8.8) is based on homomorphic unsharp masking (HUM) (Axel et al., 1987; Brinkmann et al., 1998b). HUM is a simple yet efficient approach and does not rely on any parametric representation of bias field. One of the main issues with this filter, however, is that it tends to underestimate the bias at tissue/background boundaries. There is no information in the background and still the filter tries to estimate the bias field based on neighborhood information. This in turn introduces a new illumination artifact at the tissue/background boundary. Guillemaud (Guillemaud, 1998) utilized normalized convolution to resolve this issue. Let $M(X)$ be 1 in the tissue region and 0 in the background then the bias field can be estimated as

$$B(X) = \exp \left(\frac{h(X) * (\log(I_{measured}(X)) \times M(X))}{h(X) * M(X)} \right) \quad 8.9$$

where $h(X)$ is a low-pass filter in the frequency domain. The bias field is estimated only on the tissue region and then extrapolated to the background. The Guillemaud filter works well to resolve the issue with HUM.

The low-pass filter is defined by the cut-off frequency. A very low cut-off will have no effect on the resulting image, while a very high cut-off might have adverse effects on the image such as removal of useful information. Choosing the right cut-off might not be so trivial. Cho et al highlighted that it is difficult to choose the optimal cut-off since any specific cut-off will induce different segmentation results among a set of images (Cho et al., 2004).

In this paper, we propose a new bottom-up approach to determine the optimal cut-off. As a first step, the cartilages are segmented without any NUC. Next, an intensity based measure (since NUC will induce intensity changes) known to statistically separate the two groups is quantified for all the images, for example using a t-test. In our case, this measure is cartilage homogeneity. The cut-off threshold is then varied over a pre-defined range and for each value a statistical significance level is computed. We compute the required sample size for separating the two groups. The level at which the sample size is minimum is taken to be the cut-off. The images are then corrected for based on the determined cut-off level and re-segmented. Figure 8.2 illustrates example results obtained from the method. This method will ensure that a certain cut-off is not biased towards a specific image and at the

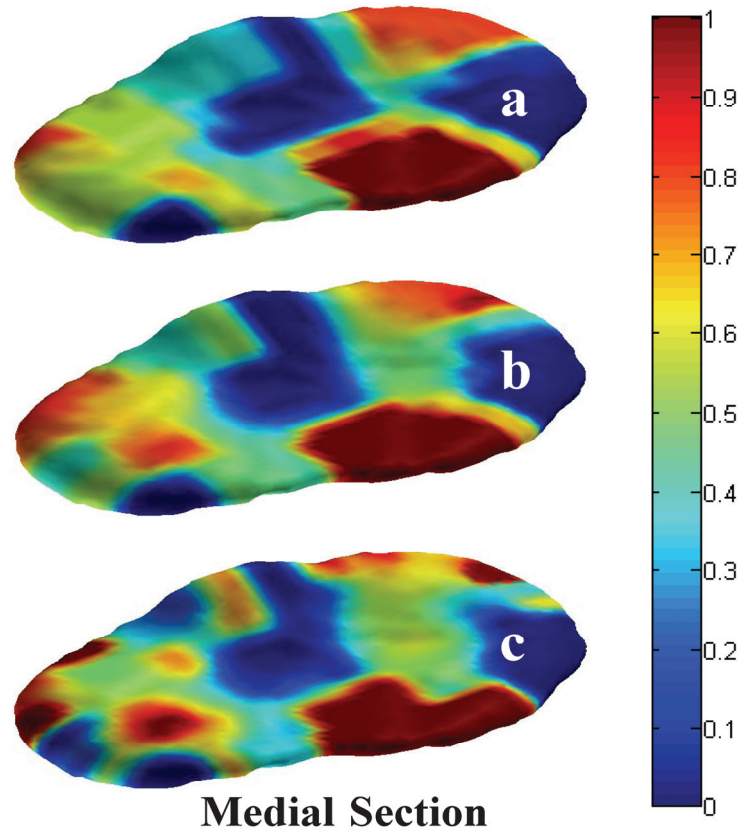


Figure 8.3: Weight maps estimated with different sub-region sizes. The measure used was cartilage homogeneity. The resolution and sample sizes are: (a) 5×10 , $N=41$. (b) 10×20 , $N=40$. (c) 20×40 , $N=38$. For visualization, the weights have been scaled to be between 0 - 1.

same time will not lead to loss of information in the filtered image, since it has been determined directly from the data.

In our implementation, K-means clustering was used to generate the M map, and Butterworth filter was chosen as the low-pass filter. The method was also evaluated on a set of brain images created by adding known biases to the Montreal Brain Phantom with no added noise (Cocosco et al., 1997).

8.5.5 Evaluation

The 286 knees were divided randomly in 3 sets consisting of 95, 95 and 96 knees respectively, and processed by Algorithm 1. The knees were divided in two groups, based on the KL index: healthy ($KL = 0$) and diseased ($KL > 0$). The sub-region resolution was chosen to be 10×20 . Experimentation with different resolutions revealed that both the resultant ROD and sample sizes were not sensitive to choice of the resolution, unless there are very few regions. Figure 8.3 illustrates the weight maps obtained at different resolutions in a sample cartilage. The resulting sample sizes are quite similar.

Table 8.2: Evaluation on Clinical Data (median sample sizes for 50 random trials)*

Measure	Set 1 (UWM)	Set 1 (ROD)	Set 2 (UWM)	Set 2 (ROD)	Set 3 (UWM)	Set 3 (ROD)	Set 3 (p-values)	Set3 (% reduction)
Homogeneity	75	27	77	50	75	61	0.04	19
Volume	228	59	187	98	189	138	0.01	36
Thickness								
RISSO	309	45	266	96	267	133	0.000001	51
LDA	309	13	266	2021	267	1927	0.0000001	-621
R-LDA	309	25	266	119	267	214	0.3	20

*UWM is the sample size, computed from a weight map with equal weights for all features. R-LDA refers to regularized LDA.

Thus, at a lower resolution we can reduce the computation time significantly without affecting accuracy. We found the resolution 10×20 to be a good trade-off between speed and accurate determination of the ROD.

Considering the biological variation of the sets and to have robust estimate of the ROD, 50 runs of Algorithm 1 were executed, each time with a different randomization of the sets. The weight maps were averaged to result in the most pathological ROD in the knee cartilage. These steps were repeated for each of the 3 measures.

Table 8.2 lists the results obtained. The median sample sizes both computed from the UWM and from the ROD, for the 3 sets are listed. The median is reported, since being less affected by outliers, it is a more stable measure, than the mean. The average reduction in sample sizes for the validation set, set 3, was: cartilage homogeneity (19%), cartilage volume (36%) and cartilage thickness (51%). In order to quantify whether the reduction in sample size was significant for set 3, we utilized the Wilcoxon Rank Sum test to compute p-values, listed in Table 8.2, testing the null hypothesis that there is no difference between the sample size from the UWM and the sample size from the ROD. Figure 8.4 illustrates the RODs in a sample cartilage.

Table 8.3: Reproducibility Evaluation of the Method

Measure	UWM (CV %)	ROD (CV %)
Homogeneity	1.6	1.9
Volume	8.6	11.1
Thickness	3.2	4.3

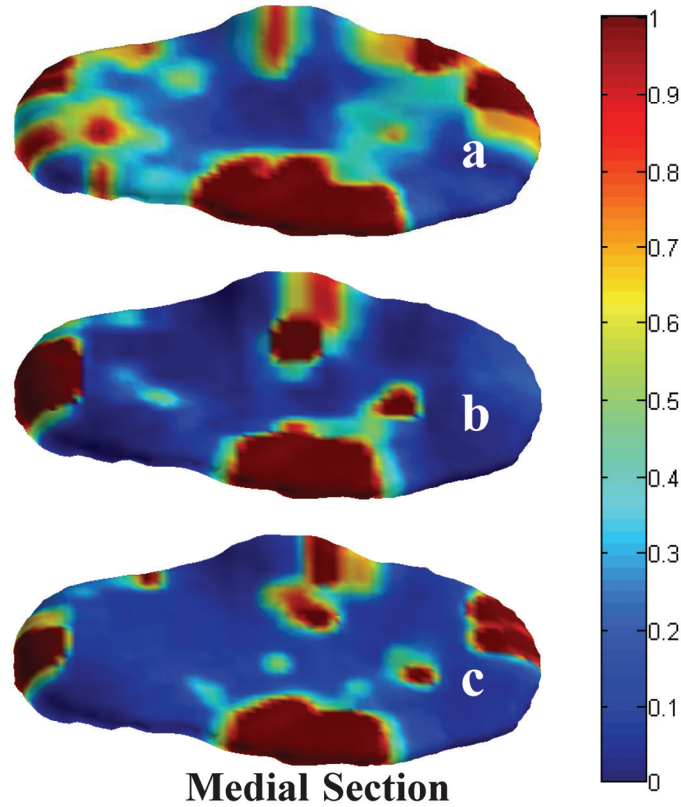


Figure 8.4: The optimal ROD projected in a sample cartilage. (a) Cartilage Homogeneity. (b) Cartilage Volume. (c) Cartilage Thickness.

Table 8.3 lists the reproducibility results for the different measures, quantified from the UWM and the ROD. The reproducibility of the ROD was assessed using the test-retest root mean squared Coefficient of Variation (RMS-CV%).

Figure 8.5, shows the mean and standard error of the mean (SEM) of the three measures for the different KL sub-groups. The figure compares the performance of the 3 measures, quantified from the UWM and the ROD.

8.5.6 Comparison with LDA

MRI studies often use cartilage thickness, as the marker for assessing cartilage damage in OA (Eckstein et al., 2006). Therefore, LDA was applied and compared to measurements of cartilage thickness. Table 8.2 lists the results obtained. The traditional LDA fails, with a 621% increase in sample size for set 3. The results improve for the regularized LDA, however, our method still performs much better with a 51% decrease in sample size for set 3, as compared to 20% by regularized LDA.

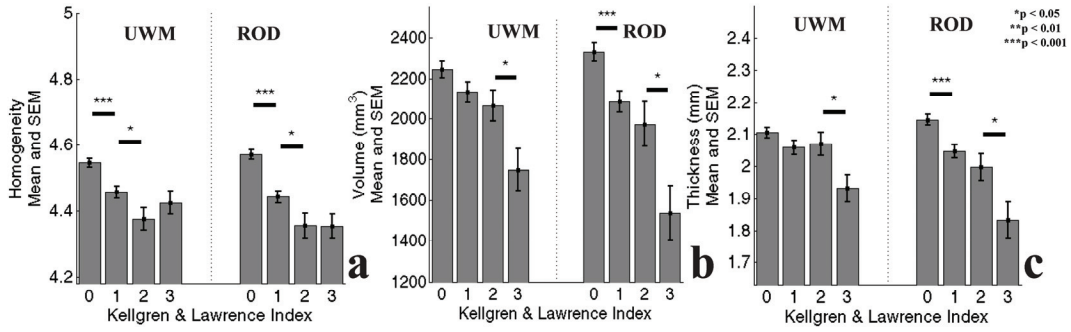


Figure 8.5: Comparing performance of the 3 measures quantified from UWM and the ROD. (a) Cartilage Homogeneity. (b) Cartilage Volume. (c) Cartilage Thickness.

8.6 Discussion and Conclusions

In this paper, we present a novel method, based on minimizing sample size to map out regions of differences between two groups of objects. We have validated the methodology on both synthetic and clinical data.

On synthetic data, we have demonstrated that our method is able to generalize to unseen data (Table 8.1), and is also able to reconstruct the ground truth ROD (Figure 8.1d). Additionally, we have compared the results to those obtained from the LDA. These results (Table 8.1) show that when compared to the LDA, our method is more robust and is able to perform reasonably well, specifically for data, which is corrupted with high levels of noise.

We have demonstrated that our method is able to localize regions in the articular cartilage using both morphometric and textural based measures (Figure 8.4). We show that the resultant regions are able to generalize (Table 8.2), lead to significant reductions in sample size (Table 8.2), and are reproducible (Table 8.3). The reduction in sample size is highly desirable for clinical studies since it will have a positive impact on the cost of the study. Additionally, the measures from the ROD are able to achieve a better quantification of the progression of the disease (Figure 8.5). For instance, for the UWM, both volume (Figure 8.5b) and thickness (Figure 8.5c) measures are not able to statistically separate the two groups of KL 0 and KL 1. The measurements, computed from the ROD, however, lead to a statistically significant ($p < 0.001$) separation. Similarly, the progression of the measures, computed from the ROD, with respect to the KL scores (specifically for thickness), follow a more natural, decreasing pattern as the disease progresses.

We have also shown that the results obtained from the evaluation of our method on clinical data are better than the results obtained from the LDA (Table 8.2). The increased performance of our method may be attributed to

the regularization terms, specifically the spatial regularization that is integrated to be a part of functional (8.7).

A limitation of the study is that the sensitivity of the method to mis-registration is not evaluated. Imperfect registration might lead to false estimations of atrophy, in case of structural features. However, our cross-validation evaluation would only improve from better alignment.

The anatomical relevance of the ROD requires insight into cartilage degeneration. Previous studies have shown that during OA the cartilage is not affected uniformly and it is believed that cartilage lesions located in the central weight-bearing region of the medial compartment are prone to more rapid cartilage loss (Biswal et al., 2002). The RODs (Figure 8.4) are not corresponding to the central, load-bearing part of the cartilage for any of the measures. Rather they predominantly outline the peripheral part of the cartilage. We believe that the region outlined is partly a result of biomechanical factors affecting the cartilage.

One such factor that has found to play an important role in OA progression is knee alignment, calculated as the hip-knee-ankle angle. Malalignment of the knee influences the neutral collinear state by shifting the load-bearing axis which in turn affects the load distribution of the knee. Previously, it has been shown by (Sharma et al., 2001) that the mechanical affects of alignment on load distribution eventually leads to OA progression in different compartments of the knee. Results from our study highlight the importance of mechanical factors in the etiology of the disease.

Recently, (Williams et al., 2006) showed that reproducibility of cartilage thickness can be increased by excluding the edges of the cartilage sheet. In contrast, our study highlights a most significant region towards the periphery. Therefore, increasing the precision by trimming the boundary of the cartilage may potentially lead to measurements that are less sensitive to the disease pathology.

The ROD based on cartilage homogeneity differs from those based on cartilage volume and cartilage thickness. In contrast to the two measures of cartilage quantity, cartilage homogeneity is a measure of cartilage quality. Thus, our framework is able to differentiate focal changes undergoing in the different phases of the disease, as highlighted in (Qvist et al., 2008).

In the future, we plan to evaluate our method on longitudinal data of the knee. This means that the computed ROD will present the regions of most changes over time. This may aid in understanding the pattern of biological changes undergoing in the articular cartilage over the course of time.